This is a digital copy of a book that was preserved for generations on library shelves before it was carefully scanned by Google as part of a project to make the world's books discoverable online.

It has survived long enough for the copyright to expire and the book to enter the public domain. A public domain book is one that was never subject to copyright or whose legal copyright term has expired. Whether a book is in the public domain may vary country to country. Public domain books are our gateways to the past, representing a wealth of history, culture and knowledge that's often difficult to discover.

Marks, notations and other marginalia present in the original volume will appear in this file - a reminder of this book's long journey from the publisher to a library and finally to you.

Usage guidelines

Google is proud to partner with libraries to digitize public domain materials and make them widely accessible. Public domain books belong to the public and we are merely their custodians. Nevertheless, this work is expensive, so in order to keep providing this resource, we have taken steps to prevent abuse by commercial parties, including placing technical restrictions on automated querying.

We also ask that you:

+ *Make non-commercial use of the files* We designed Google Book Search for use by individuals, and we request that you use these files for personal, non-commercial purposes.

+ *Refrain from automated querying* Do not send automated queries of any sort to Google's system: If you are conducting research on machine translation, optical character recognition or other areas where access to a large amount of text is helpful, please contact us. We encourage the use of public domain materials for these purposes and may be able to help.

+ *Maintain attribution* The Google "watermark" you see on each file is essential for informing people about this project and helping them find additional materials through Google Book Search. Please do not remove it.

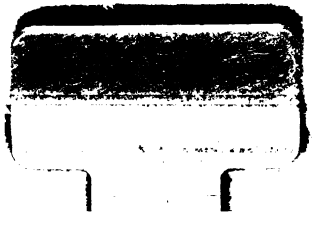
+ *Keep it legal* Whatever your use, remember that you are responsible for ensuring that what you are doing is legal. Do not assume that just because we believe a book is in the public domain for users in the United States, that the work is also in the public domain for users in other countries. Whether a book is still in copyright varies from country to country, and we can't offer guidance on whether any specific use of any specific book is allowed. Please do not assume that a book's appearance in Google Book Search means it can be used in any manner anywhere in the world. Copyright infringement liability can be quite severe.

About Google Book Search

Google's mission is to organize the world's information and to make it universally accessible and useful. Google Book Search helps readers discover the world's books while helping authors and publishers reach new audiences. You can search through the full text of this book on the web at `http://books.google.com/`

UC-NRLF
$B 17 848

Digitized by Google



Digitized by Google

Digitized by Google

Digitized by Google

Trade Tests in Education

By

HERBERT ANDERSON TOOPS

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy, in the Faculty of Philosophy, Columbia University

Published by

Teachers College, Columbia University

New York City

1921

Digitized by Google

Copyright, 1921, by HERBERT ANDERSON TOOPS

UNIV. OF CALIFORNIA

Digitized by Google

G. E.

ACKNOWLEDGMENTS

In a work of this kind, reporting as it does in large part the methods developed by an army group, the author is evidently indebted to many persons, both directly and indirectly. To Dr. E. L. Thorndike, whose direction and inspiration has been such as to lead to a greater undertaking in this work than originally planned, much special credit is due for such merit as this work may possess. Special mention should be made of others whose guidance and aid have directly been very helpful in this work: Dr. Truman L. Kelly, whose statistical guidance has made possible the analysis methods developed herein; Dr. Arthur D. Dean, whose vocational philosophy is reflected from many pages of this book; Miss Lucy L. Brown, teacher in charge of the vocational testing classes at the Manhattan Trade School for Girls, whose coöperation and interest have made possible the analysis of vocational guidance data; Dr. A. H. Ryan, director of the Industrial Hygiene Department of The Scovill Manufacturing Co., for industrial data bearing on the evaluation of talents for the job of youths newly entering industry; Dr. Howard G. Burgess, of the Vocational Training Department of the New York Military Training Commission, who has supplied statistics on the grades completed in school by working boys recently leaving school in the state in New York; Mr. James D. Jackson who spent many weary hours in most helpful statistical aid; Drs. J. Crosby Chapman, William McCall, and Arthur S. Otis who gave valuable aid in special aspects of this work. Acknowledgment is made to a host of others, whose contributions, although not specially mentioned, are none the less valuable.

HERBERT A. TOOPS

456685

Digitized by Google

Digitized by Google

CONTENTS

CHAPTER PAGE

I. THE VALUE AND USES OF TRADE TESTS 1

1. The value of trade school instruction a debatable issue.
2. Lack of units in which to measure trade proficiency.
3. The value of trade tests in vocational education.
   - A. In testing progress, and maintaining interest in trade instruction.
   - B. In making proper allowances for individual differences in rates of acquiring trade proficiency.
   - C. In educational and vocational guidance and school placement work.
4. The broadness of scope of "trade test methods."
5. The classification of trade tests.
   - A. Tests of proficiency or success in a trade.
     - *a*. Description of the oral, picture, and performance trade test methods, and reproduction of a typical test of each.
   - B. Tests of trade capacity, potentiality or promise.
     - *a*. The varied nature of tests of trade capacity.
     - *b*. Description of a typical performance test of ability to learn a new trade operation.
   - C. The fundamental basis of all trade tests, a statistical technique common to all test methods.

II. TESTS OF TRADE PROFICIENCY OR SUCCESS, AND THEIR ADAPTATION TO SCHOOL WORK 17

1. Comparison of the tests of the interview and "hiring on trial" with the oral trade test method.
2. The technique of trade test construction.
   - A. Selection and revision of oral trade test questions.
   - B. Selection and revision of picture trade test questions.
   - C. Standardization of trade tests.
   - D. Selection of tasks for performance trade tests.
   - E. The multiple choice written trade test.
     - *a*. An experiment, with results obtained in the use of a written bricklayers' and a farmers' test.
   - F. An experiment, with the results obtained from a written information test made up in three different examination methods.
3. The factors determinative of the best examination method for school use.
   - A. Adaptation of the one-word-answer question to trade and non-trade school examinations.
   - B. The merits and possible criticisms of trade test methods in school examinations.

Digitized by Google

III. TESTS OF TRADE CAPACITY AND RESULTS OF THE EVALUATION OF THEIR USE IN A TRADE SCHOOL . . . 63

1. The evolution of present-day test philosophy.
2. The problem of industrial placement.
   A. Results from tests of vocational capacity secured by other investigators.
   B. Use of job analysis in vocational placement.
3. An experiment, with results from the evaluation of trade school tests of trade capacity.
   A. The testing department.
   B. The tests.
   C. Methods of, and results from evaluation of qualification card and supplementary data.
      a. Intercorrelations of the tests.
      b. Follow-up report on tested pupils who later entered commercial high school.
      c. The value of the tests.

IV. GENERAL CONSIDERATIONS IN TESTING AND TRAINING FOR PROFICIENCY AND PROMISE IN THE TRADES . . 96

1. Trade school graduates not the tradesmen of industry.
2. Specialization in industry and its effects upon tradesmen's talents for the job.
3. The intellectual narrowness of industrial trade training.
4. Distributions of talents of tradesmen, and their significance for training and selection of tradesmen.
   A. Tradesmen's intelligence.
   B. Tradesmen's general education.
5. The contribution of general education to vocational proficiency.
   A. Results in the case of general clerks.
   B. Results in the case of eyelet machine operators.
6. Changes in trade school administration suggested by trade test methods.
   A. Adaptation of trade school instruction to the intelligence level of the pupils.
   B. Use of trade examinations as learning incentives.
   C. The standard job method.
   D. The merits of the self-administrative recall examination.

APPENDIX: STATISTICAL METHODS 112

BIBLIOGRAPHY 116

Digitized by Google

CHAPTER I

THE VALUE AND USES OF TRADE TESTS

Vocational schools, as opposed to schools for general education, have as their avowed aim the preparation of youth to earn a living in vocations.

The extent to which their aim is achieved by such schools as are called vocational schools is at present a debatable issue. One reason for this is the lack of means with which to measure the human products of such schools. Inability to know the product of the school at once predetermines an inability to better the school administration on a basis of the needs of students. Any betterment in present methods of measuring the product of a school at once produces the chance for bettering the school administration on the basis of the needs of the students. Only until very recently has trade skill and knowledge been measured in more refined terms than subjective personal judgments of "skilled man" and "helper," "successful" and "unsuccessful," "competent" and "incompetent." Army trade tests solved this problem for the army; adaptations of the method may solve the problem for our vocational schools. Graduates of our vocational schools are at present of equal merit, so far as measure of their merit goes, for they all receive diplomas. Industry wants a more accurate measure of their hiring worth; and the school desires to recognize the varying merit of its product.

Unless the prospective employer can successfully rate the human product of the vocational school he is at a loss to know in placing a graduate of a vocational school in industry whether he should adopt a course different from that followed in the case of any other applicant for the job. The average employer is skeptical of the value of the "book learned" tradesman. His skepticism, we must believe, is not directed so much toward vocational school graduates as against particular exceptional individuals in the past who have happened to fail when given the test of industry, the test of doing the job itself. Could graduating students be rated for general trade proficiency against men in the trade itself, much would be gained in the way of proving just what is the value of any specific vocational training program.

Digitized by Google

UNIV. OF CALIFORNIA

The percentage of a graduating vocational school group which equals or exceeds in general trade skill the median tradesman in industry is a theoretical measure of the efficiency of the vocational school instruction. The army trade tests, standardized on random sampling of men in industry, fulfill these requirements of measures of final trade proficiency after a period of training. They may be used equally well to measure trade proficiency at any point of its acquirement.

Vocational school instruction has so generally meant trade school instruction that the author will not attempt here to make a distinction between the two. It is evident that there is no distinction between the two in the field of tests, which have a common statistical technique and hence make no distinction, save in content, between intelligence tests, educational tests, vocational tests, trade tests, or any other variety of test which has as its aim the prediction of peoples' abilities to do something, or to act in a certain way. Accordingly, the terms "trade test" and "vocational test" will be used interchangeably.

VALUE OF TRADE TESTS IN EDUCATIONAL ADMINISTRATION

The values of trade tests in education are such as pertain to any objective, standardized trade examination method. Standardized tests, or tests designed by taking into account the testing principles thus far discovered, will have a better chance of fulfilling the demands made upon them than tests not so constructed.

For argument's sake, one may take for granted the assumption that a vocational education course under consideration is a desirable course, and that it is as good a course as may be expected to be produced. Not all students will apply themselves equally to the tasks of learning the trade instructions given them in such a course, nor profit equally from those received. Tests of progress, given at periodical intervals, are needed to keep pupils up to a reasonable standard of interest in their courses and for purposes of grading students at the end of the year, or at the end of the course. If such tests can be standardized, then so much the better. But if not, as will often be the case, then vocational instructors may be able to profit greatly by attention to the methods of question making and test formulation used in army procedure.

Digitized by Google

Of two students of equal general trade or vocational ability, one may be unusually good in operation A and be unusually poor in B; the other unusually poor in A but good in B. If, during the course of instruction, the strengths and weaknesses of each student could be made known, much valuable time of the student might be saved, and a much better allround ability gained by the teacher's adapting his teaching to individual needs. Diagnostic tests of analyzed or subdivided elements of general trade ability are needed. These will be developed in the future. A knowledge of what the army was able to do in measuring trade ability should help the prospective test maker in making up such tests. Where such tests have been developed to the point where they are sufficiently accurate for the purpose, we may expect trade instruction to be done largely by the job or operation sheet method, and the pupil to be advanced from job to job or operation to operation just as fast as he shows a test measured proficiency of *X* in each job or operation. Thus will result economy of time required in learning a trade, since each pupil will progress just as fast as he is capable of doing. This is the method found successful by the S. A. T. C. during the war, and since adopted by the E. and R. Schools of the Army in teaching trades to thousands of boys. In the past the test of whether or not a given pupil was ready for promotion to a new job has been a subjective estimate by the instructor, aided by performance on the job, it is true, but nevertheless a test which could not be duplicated by an outside person without a personal knowledge of the pupils being rated.

Tests of promise or capacity have been used in the field of general education for a number of years; they have not been so generally used in trade education. Much research work, greatly encouraged by the success of the army intelligence tests, has demonstrated the feasibility of classifying school children on the basis of their mental ability, or on the basis of a composite score of a number of tests designed to predict pupils' abilities to profit by the instruction given them. The E. and R. Schools of the Army have thus classified many thousands of soldiers, drawn from every part of the country and from foreign countries, with varied and non-comparable previous school experiences. The value of such educational guidance has been already proved.

Trade test methods, making use of refined tests of trade capacity,

Digitized by Google

promise to throw some light upon the age-old problem of "What vocation should a boy enter?" Whether a boy should take up general education, trade education, or professional education, is a problem which we may hope to solve only by development of a vocational philosophy formulated with the aid of pertinent facts of human abilities and vocational opportunities made available by the statistical methods common to all survey methods. In conjunction with psychological tests, physical tests and questionnaires, tests of trade capacity may be expected to become a valuable aid to the vocational counsellor of the future.

The best large vocational schools have their own placement or employment bureaus. The problem of such bureaus is to evaluate into one single fitness score, by a subjective process, the diversified talents of boys and girls as a basis for a twofold categorical division of resulting action into "recommended" and "not recommended" for the prospective position, whether job in industry, or school instruction in an advanced school. Such estimates, when based upon proficiency in performing trade operations, may be greatly enhanced by more accurate methods of measuring trade skill. The two aspects of such placement are: (*a*) knowledge of the pupil's abilities, and (*b*) the requirements of the job, in the way of human abilities and acquirements. Text-books on vocational guidance give many facts about trades, tradesmen and trade conditions. They generally fail to specify how diverse human talents for the job, or diverse demands of the job upon the worker, may be evaluated. These problems will be ultimately solved by methods available to, or to be developed by, the test statistician.

This larger aspect of "fitting the job and the man" is even more important than the more restricted problem of measuring the talents for a job, present or future, measurable by trade tests. All make use of a common statistical technique. Accordingly, we should mean by "trade test methods" this larger aspect of evaluating measurable human talents in terms of some independent criterion of job proficiency. Inasmuch as tests of present trade proficiency are promises of future performances (immediate or remote) on the job, all trade tests in the last analysis are tests of promise or capacity. The common statistical problem of all such methods is the problem of measuring mutual dependence of related measures of trade capacity.

Digitized by Google

CLASSIFICATION OF TRADE TESTS AND EXAMPLES OF TYPICAL FORMS

The term "trade test" is used to refer to any test involving test material taken from the trades. Trade tests differ, then, from any other forms of tests mainly in a difference in content. They may be categorically divided into two classes according to the purpose of the test, and further sub-classified according to form of test material and method of administration, as follows:

I. Tests of proficiency or success in a trade.
   1. Verbal.
      a. Oral Administration.
         (1) Army Oral.
            (*a*) Oral Answers.
            (*b*) Written Answers.
         (2) Picture.
      b. Written Administration.
         (1) One-word-answer.
         (2) Multiple Choice.
   2. Performance Tests.

II. Tests of trade capacity, potentiality, or promise.

Sub-classes of this group: Practically every known variety of test, including the above verbal forms, has been used by some investigator at one time or another.

Trade tests were used in the army to determine which men, of those claiming specific trade experience, were sufficiently competent to be detailed for special duty in the trades maintained in the army.

An example of an army oral form of test, used for measuring trade proficiency, is reproduced herewith in an eyelet machine operator's test. This test may be given individually and orally by an examiner, the answers being oral; it may be given orally to a group, the answers being written on blank paper; or it may be administered by the use of printed or mimeographed questions, the answers being in the one-word-answer form.

EYELET MACHINE OPERATOR (TENDER)

(*Army oral form of test with answer in the one-word-answer form.*)

1. What do you call the large main shaft at the top of the machine?
   Ans. Cam.
2. What rubs on the cams all the time to oil them?
   Ans. Tassel.

Digitized by Google

3. What does the scrap wind up on?
   Ans. Reel.
4. What drives the reel from the bottom cam shaft?
   Ans. Belt.
5. What do you call the die which does the first operation?
   Ans. Blanking.
6. What is the name of the large part which moves back and forth and carries the work from one die to another?
   Ans. Slide.
7. What do you call the part which moves up and down and carries the punch at its lower end?
   Ans. Plunger.
8. What do you call the pulley on the drive shaft which is not fastened to the shaft?
   Ans. Loose.
9. What is the shaft underneath the machine called?
   Ans. Cam.
10. What are the small springlike parts called which grab the work and carry it from one die to another?
    Ans. Finger.
11. What do you use to measure the thickness of the metal?
    Ans. Micrometer.
12. What do you use to cut off the strip metal?
    Ans. Shears (snips).
13. What is used to tell if the finished work is the correct size?
    Ans. Gage.
14. What do you call a gage which shows the largest and smallest size that can be allowed on a finished piece of work?
    Ans. Limit (maximum and minimum) ("max and min").
15. Of what metal are the reels made?
    Ans. Brass.
16. What is the shape of the plungers?
    Ans. Square.
17. When five rows of blanks are to be punched from the strip metal, which row is punched first?
    Ans. Middle (center).
18. What is there on the end of the air pipe to let the air out at just the right time?
    Ans. Valve.
19. How many holes are there in the drive pulley to turn it by hand?
    Ans. 2 (4).
20. What do you call the tool used to turn the fly wheel by hand?
    Ans. Bar.
21. What part of the cam shaft regulates the air?
    Ans. Cam.
22. If the top cam shaft turns at 60 revolutions per minute how many revolutions per minute does the bottom cam shaft turn?
    Ans. 60 (same).

Digitized by Google

23. What guides the finished work into the pan?
Ans. Spout (pipe).
24. What is the sheet iron around the gears called?
Ans. Guard.
25. What does the pan for the work set on?
Ans. Box.
26. What do the dies set in?
Ans. Holder.
27. What tool do you use to turn the roll feed?
Ans. Wrench.
28. What material is used to line a brake on machines which have a brake on the drive shaft?
Ans. Leather (wood).
29. How is the punch fastened to the plunger?
Ans. Screws in (wrench) (thread).
30. What is the washer made of that is sometimes used on the reel shaft?
Ans. Leather.
31. What do you call the attachment sometimes used around a punch to make the work let loose from the punch?
Ans. Thimble.
32. How are the bottom cams made so that they can be timed?
Ans. Split (halves) (screw).
33. What tool do you use to set the bottom cams on the bottom cam shaft?
Ans. Screw-driver.
34. Of what material is the part of the slide friction made that rubs on the slide?
Ans. Wood (maple).

Army picture tests were administered orally and the answers given orally. They were used as a check on the ratings made on the oral tests, which are somewhat subject to language difficulties and to coaching. The method is essentially the army oral method, but with pictures used to describe the trade situation in the place of the descriptive phrases or clauses which must be used in the oral method. The lathe hand test, herewith reproduced, pages 8 to 14, is typical of this form of examination.

The performance test aims to duplicate a shop situation, the test subject being rated on the quantity and quality of work done on a job or problem assigned. These have the merit, not possessed by ordinary shop jobs, of being so chosen as to require a maximum amount of manipulation of tools, use of trade knowledge and trade skill. A sample performance test, used in testing ability to use machinists' measuring tools, is shown on page 15. This particular test has as its products a series of dimensions entered on a form blank, which are scorable by stencil.

Digitized by Google

Tests of trade proficiency were developed later than tests of trade capacity. Tests of trade proficiency were developed to a high point of efficiency during the War. The possibilities of large scale research in developing tests of trade capacity have not yet been tested. It is the special function of this research to point out the value of such work by statistical inquiry into particular selected vocations and trade school methods.

LATHE HAND TEST

PICTURE 1

1. What is the part at G called?
   Ans. Steady-rest.
2. What letter shows the live center?
   Ans. D.
3. What is the lever at K used for?
   Ans. Screws (threads).
4. What do you call the part at C?
   Ans. Face-plate.
5. What do you call the lever at A?
   Ans. Back-gear.
6. What is the part at L used for?
   Ans. Gears (change the gears).
7. What letter shows the cone?
   Ans. B.
8. What letter marks the tail stock?
   Ans. H.
9. What is the part at F called?
   Ans. Follower-rest (follower).
10. What is the part at J called?
    Ans. Rack.
11. What is the part at I called?
    Ans. Lead-screw.
12. What is the part at E called?
    Ans. Tool-post.

PICTURE 2

13. What does that micrometer read?
    Ans. .125.
14. What is the part at A called?
    Ans. Ratchet.

PICTURE 3

15. What do you call that lathe tool?
    Ans. Dog.
16. What does the part at A fit into on the lathe?
    Ans. Face-plate (slot on face-plate).

Digitized by Google

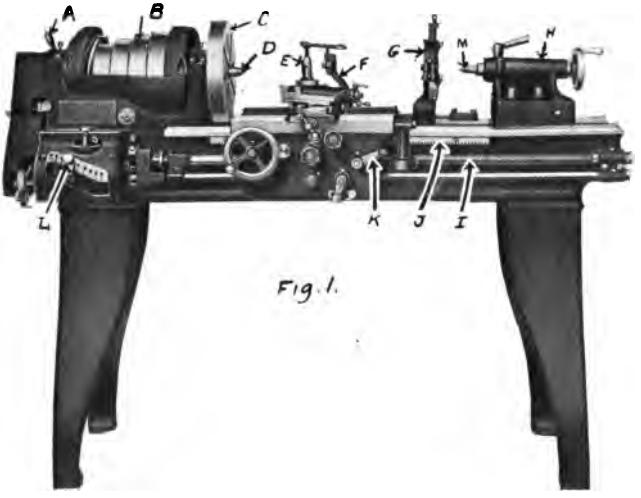


Fig. 1.

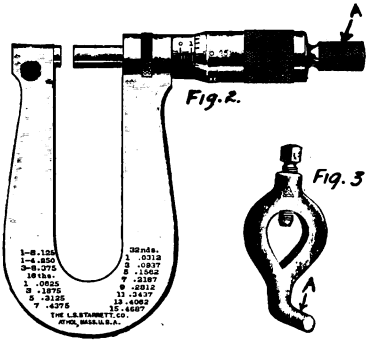


Fig. 2.

Fig. 3

Fig. 4.

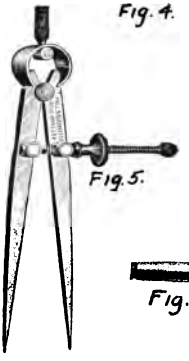


Fig. 5.

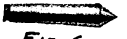


Fig. 6.

Fig. 7.

Fig. 8

Fig. 9

PLATE I. PAGE 1 OF LATHE HAND PICTURE TRADE TEST

Digitized by Google

PICTURE 4

17. What do you call that tool?
Ans. Boring.

PICTURE 5

18. What is that tool called?
Ans. Dividers.

PICTURE 6

19. What do you call that lathe tool?
Ans. Center.
20. How many degrees in the angle of the point of that center?
Ans. 60.

PICTURE 7

21. What do you call that tool?
Ans. Center-punch.

PICTURE 8

22. What is that tool used for?
Ans. Threads (screws).

PICTURE 9

23. In what operation is that tool used?
Ans. Drilling.

PICTURE 10

24. What letter marks the apron of that lathe?
Ans. G.
25. What letter marks the bed of that lathe?
Ans. H.
26. What letter shows the part which you turn to throw the longitudinal feed in gear?
Ans. C.
27. What letter shows the part which you turn to throw the automatic cross-feed in gear?
Ans. D.
28. What letter marks the part which works the compound-feed?
Ans. I.
29. What letter shows the swing of that lathe?
Ans. A.

PICTURE 11

30. What operation is being done in that picture?
Ans. Boring.
31. What is the part at A called?
Ans. Face-plate.
32. What is the part at B called?
Ans. Dog.

Digitized by Google

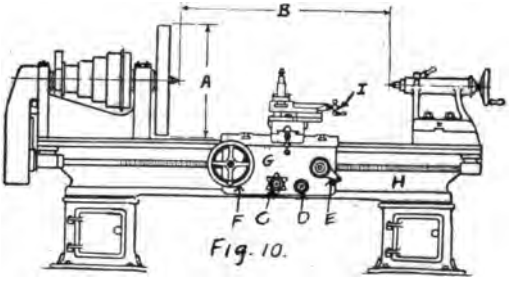


Fig. 10.

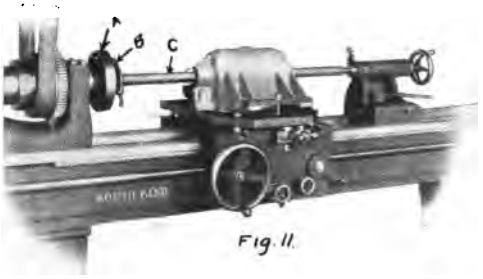


Fig. 11.

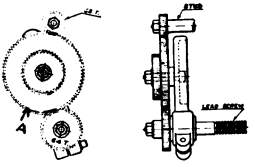


Fig. 12.

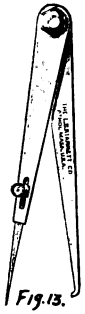


Fig. 13.

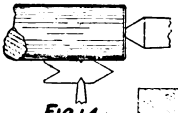


Fig. 14.

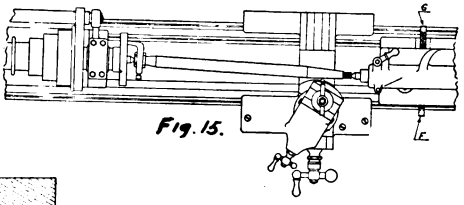


Fig. 15.

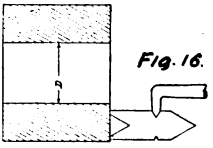


Fig. 16.

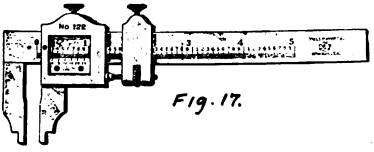


Fig. 17.

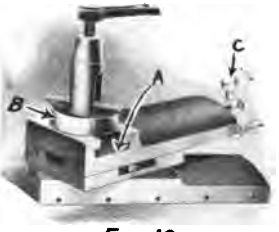


Fig. 18.

PLATE II. PAGE 2 OF LATHE HAND PICTURE TRADE TEST

Digitized by Google

33. What is the part at C called?
Ans. Boring-bar.

PICTURE 12

34. If the lead screw on that lathe has 8-threads per inch, how many threads per inch will that lathe cut with those gears?
Ans. 16.

35. What is the gear at A called?
Ans. Idler (intermediate).

PICTURE 13

36. What is the tool in that picture called?
Ans. Morphadite (hermaphrodite).

PICTURE 14

37. What tool is being squared-up in that picture?
Ans. Threading.

PICTURE 15

38. What operation is being done in that picture?
Ans. Taper.

39. If the stock in that lathe is 2-feet long and you want a taper of ¼-inch per foot, how far do you set-over the screw G?
Ans. ¼ (¼-inch).

PICTURE 16

40. What tool is being squared-up in that picture?
Ans. Inside-threading.

PICTURE 17

41. What do you call that tool?
Ans. Vernier-caliper.

42. What is the smallest fraction of an inch which that will measure?
Ans. 1/1000.

PICTURE 18

43. What do you call the part at B?
Ans. Collet (collar) (washer).

44. What do you call a slot like that at A?
Ans. T-slot.

45. How far does one revolution of the handle at C advance the point of the thread tool?
Ans. ⅛ (⅛-inch) (.125).

PICTURE 19

46. What is the name of the tool at 1?
Ans. Left-hand-facing.

47. What is the name of the tool at 2?
Ans. Right-hand-facing.

48. What is the name of the tool at 4?
Ans. Diamond-point.

Digitized by Google

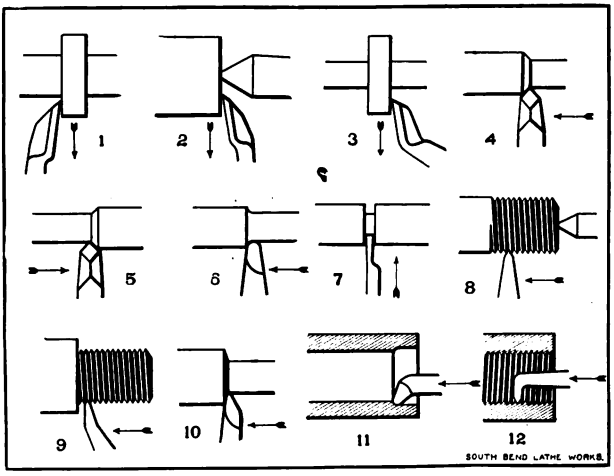


Fig 19.

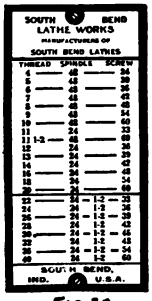


Fig.20.

Fig.21.

Fig.22.

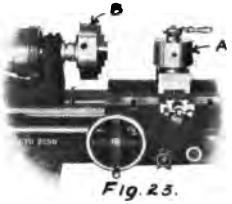


Fig.23.

Fig.25.

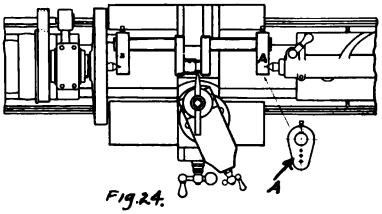


Fig.24.

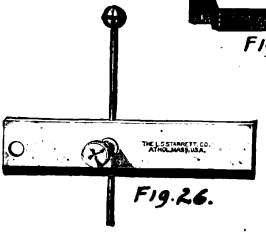


Fig.26.

PLATE III. PAGE 3 OF LATHE HAND PICTURE TRADE TEST

Digitized by Google

49. What is the name of the tool at 6?
Ans. Round-nose.
50. What is the name of the tool at 7?
Ans. Parting (cutting-off).
51. What is the name of the tool at 9?
Ans. Bent-threading.
52. What is the name of the tool at 12?
Ans. Inside-threading.

PICTURE 20

53. What gear do you use on the lead-screw to cut a screw of 13 threads per inch, if you have a 24-tooth gear on the spindle?
Ans. 39.
54. How many threads per inch on a 1-inch pipe thread?
Ans. 11½.

PICTURE 21

55. What kind of work has been done on that stock?
Ans. Knurling.

PICTURE 22

56. What do you call that lathe tool?
Ans. Drill-chuck.
57. What do you call the set-screws on that chuck?
Ans. Safety.

PICTURE 23

58. What do you call the part at B?
Ans. Chuck.
59. What do you call the part at A?
Ans. Turrett.

PICTURE 24

60. What is being turned-up in that picture?
Ans. Crank-shaft.
61. What is the part at A called?
Ans. Jig.
62. What do you use to prevent vibration when turning off-center work?
Ans. Weight (balance) (counter-balance).

PICTURE 25

63. What do you call that tool?
Ans. V-block.
64. What is that tool used for?
Ans. Centering (laying-out).

PICTURE 26

65. What do you call that tool?
Ans. Depth-gage.

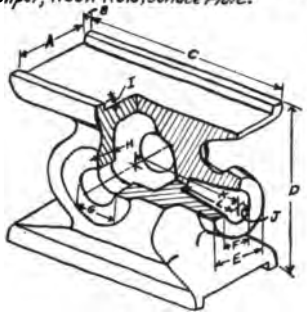
Digitized by Google

Tests of trade capacity, potentiality, or promise range widely in content, administration, and test method. At one end of the range of tests so used in the past, we have the simple tests of psychophysical measurements; at the opposite end, we have lengthy questionnaires aiming at measuring amateur ability in the trade, and initial demonstration trade lessons with performance tests thereon, designed to measure ability to learn simple processes of

Name _____ Check No. _____

Tool Maker, Lathe Hand, etc. Measurment Performance Test.

Tools Required. Machinist's Scale ($\frac{1}{100}$ and $\frac{1}{64}$), Depth Gage, Surface Gage, Taper Gage, Inside and Outside Calipers, Micrometer Caliper, Hook Rule, Surface Plate.



Directions: Measure the casting at the points indicated, and write the results on the chart.

DO NOT MARK CASTING.

Meas. No.	DESCRIPTION	Measurment in Inches (Decimals)
A	Width ±.01	
B	Width ±.001	
C	Length ±.01	
D	Height ±.01	
E	Outside Diameter ±.0005	
F	Inside Diameter ±.01	
G	Inside Diameter ±.001	
H	Thickness ±.01	
I	Height ±.01	
J	Taper per foot ±.02	
K	Centre to face dist ±.02	
L	Depth of Hole ±.001	

FIG. 27. MEASUREMENT PERFORMANCE TEST FOR TOOLMAKER, LATHE HAND, DESIGNER, ETC. DESIGNED TO INVOLVE USE OF PRINCIPAL MEASURING TOOLS. THE CASTING IS MACHINED AND POLISHED.

the trade. The apparatus used in administering such tests has varied from a piece of paper and pencil to elaborate apparatus designed to imitate the motions of the job while automatically recording quality and quantity of performance. Such tests have been used to predict probable wages, amount of production per hour, foreman's rankings of workers, accident risk, school marks, chance of being hired when referred to a job, probable length of stay on the job, and so on.

As typical of one of the most recent developments of tests of trade capacity might be mentioned a pasting test used at the Manhattan Trade School for Girls, New York City. The teacher, with all required materials before her, demonstrates the making of a paper

Digitized by Google

covered pasteboard box. Each step in the operation is explained: the use of tools, position of fingers, position of materials, chances for doing poor work which must be avoided by the pupil. At the completion of the demonstration each child takes her seat, and is provided with paste-pot, paste, brush, wipe-rag, scissors, pastingboard, pasteboard, colored paper, cloth for reënforcement of covers, and rule. Without further instruction the pupil makes the best box she can, guided by the instructions remembered. The final score on the performance is subjectively judged by the teacher for "accuracy, neatness, speed and handling," which four are finally subjectively evaluated into an "estimate of pasting ability." It will thus be seen that the test involves ability to imitate recalled directions and demonstrated trade performance; in short, ability to learn a new trade job.

It is unnecessary to emphasize the fact that, in so far as test technique is concerned, there is nothing essentially new or different in a trade test from tests used in the past, called intelligence tests, educational tests, or vocational tests. The growing use of the statistical method of partial correlation has helped to emphasize the fact that any measurable variable fact about a group of people may be used to predict any other variable fact about that same group with which it correlates either positively or negatively. Almost any verbal intelligence test will be found to be a fair predicter of school marks; school marks may, in certain vocations, be fair predicters of probable after-school success in industry; marital condition may be a fair predicter of a man's chances of being hired when referred to a prospective employer for a job. The essential common basis of all such scientifically constructed tests,—and trade tests in particular,—are: (*a*) determination of a numerical value for both variables, test score and measure of success, on a group of persons of known or judged ability, and (*b*) determination of the correspondence of test score to job ability to be predicted, on the hypothesis that the most probable ability of a second person, of unknown ability on the job, but with a given test score X, is the ability of the standardization group of persons which corresponds to the test score X. The variable, degree of success on the job, which is to be predicted, is generally known as the "criterion." Due to lack of standardization of workmen's job status in industry, a criterion of trade ability is unusually hard to establish. The statistical technique used is employed in other kinds of tests as well.

Digitized by Google

CHAPTER II

TESTS OF TRADE PROFICIENCY OR SUCCESS, AND THEIR ADAPTATION TO SCHOOL WORK

COMPARISON OF THE TESTS OF THE INTERVIEW AND "HIRING ON TRIAL" WITH THE ORAL TRADE TEST METHOD

The interview has long been the basis upon which men have been judged in regard to their fitness for jobs. When questions, specifically applying to the qualifications for the job, have been asked the applicant, these have formed one kind of trade test.

Previous to the war, little constructive thought had been given to the methods of examining tradesmen. Many studies of labor turnover had clearly indicated that the causes inherent in the way men are fitted into jobs are often the major causes of such labor turnover. Ways of bettering the selection or examining process were slow in being developed.

One of the most common methods of selection of men has been the process of "hiring on trial." The costliness of the method in time spent in supervision of the new employee, spoiled materials, broken machinery and dissatisfaction of the employee judged incompetent after trial make the method too costly for present-day use. Consequently, one may say that present-day general practice is to hire men by an interview given them at the employment office.

Some of the most apparent defects in the method will be discussed, for it is just such inferior methods of examination that any graduate of any school has to face when applying for a job.

1. The applicant is seldom given many pertinent questions about the job for which he is being examined. The routine of filling out a qualification card usually takes so much time that little time is given to the much more important consideration of the trade qualifications possessed by the applicant. Statistically, the objection is that where few questions are used, the ratio of the P.E. of an individual score to the magnitude of the score is too large.

2. Not all applicants get the same set of questions, but each applicant receives the questions which come to the examiner's mind

Digitized by Google

on the spur of the moment. Even if the evaluation of the questions were objective, comparable ratings would be impossible. Depending upon the questions asked, the difficulty of examinations received by successive applicants varies greatly.

3. Because of insufficient care in wording the questions, the questions which are asked may be generally answered satisfactorily by a person with only a very superficial knowledge of the trade process referred to. Thus questions are often merely a test of the workman's ability to understand the language of the examiner rather than a test of his trade knowledge or skill.

4. "Trick" questions have been common. The assumption underlying these seems to be a feeling that the truly expert tradesman will not be caught by the subtle trap set for him by the examiner.

5. Guess questions, or questions adequately answerable by "yes" or "no," have often been used. This form of test, commonly known as the "true-false" test, is a common form of psychological test method. In order to avoid a high P.E. of the individual score, test makers generally use a large number of such questions and employ a scoring formula which penalizes errors.

6. Owing undoubtedly to the seeming failures of the verbal method in the past, a very general belief has arisen among interviewers, and trades people in general, that the only reliable way to test a workingman's trade ability is to watch the workman at his work. Consequently, the trade examination is often slighted. Practically no attempts, outside of a few experiments by psychologists, have been made to prove statistically the worth of one examination as against any other.

7. Using the common type of question demanding long explanatory answers, much time would necessarily have to be devoted to the trade examination in order to sample but a very few of the elements of trade skill and knowledge possessed by the tradesman.

8. Owing to the time limitations and necessity of using as examiner a person well versed in the trade and in methods of oral examination, it is impossible to use the method where great numbers of persons in various trades might be desired to be examined with many pertinent trade questions. One of the greatest advantages of the personal oral examination is the use of pertinent follow-up questions, used by the examiner as a sort of trade cross-examination method. The method is not adaptable to written trade examinations,

Digitized by Google

such as are desirable, and often necessary, in all trade or vocational school examinations.

9. In any follow-up of graduates of a trade school, for instance, the only available statistical record kept of the oral interview is the note on the qualification card of "hired" or "not hired." Such categorical ratings are entirely too crude for accurate statistical studies. Furthermore, the judgment thus given is a composite subjective judgment of not only the results of the trade questions but of the qualification card entries as well. It is thus impossible to tell how much weight the one factor may have had, as against the others, in determining the final categorical rating.

10. For any reasonable approach to accuracy, the scoring of the answers given demands an expert in the trade to evaluate the tradesman's answers. In any modern large industrial employment office, men are hired in many trades and for, literally, hundreds of different jobs. It is obviously impossible for one man to be intimately acquainted with the trade processes of so many trades and jobs. The self-styled "expert" examiner, after having once discovered a few trade questions on each trade, usually lets his examinations thereafter be wholly made up of the few trade questions which come to mind on the spur of the moment. The questions are easily coached for by intelligent workmen, while the method fails utterly as do all oral methods, in the case of illiterate or foreign-born workmen.

Many of the above enumerated disadvantages of the present verbal methods of examination are overcome by the army oral examination. This examination owes its value largely to the methods used in obtaining the information which will make good trade questions, and in the methods of wording the questions. These advantages are as follows:

1. The questions are assembled in conference with experts in the trade, only those questions being selected for standardization which have been proved to be answerable by a majority of the expert tradesmen in the trade in question. The questions are thus, first of all in the assembly process, selected questions; they are not such questions as expert opinion might be convinced were desirable questions, but rather questions such as tryouts on experts in the trade have demonstrated to be pertinent questions.

2. The questions are later standardized on men in industry of known different degrees of ability, and a further selection of ques-

Digitized by Google

tions is made upon the statistical basis of proved differentiation of different degrees of trade skill on the part of each question used in the final test.

3. Due both to the clearness in meaning to the applicant of the questions and to the short key-word answers demanded by the questions, many questions may be given in a short time. When examining expert tradesmen, a speed of as many as three or even four questions given and answered per minute is possible. This means that many samplings of diverse elements of trade ability may be made in the same time as formerly devoted to but few questions.

4. The key-word scoring method makes the scoring thoroughly objective. All possible allowable answers are predetermined at the time of the compilation of the examination. Consequently, any person competent to recognize the printed answer is practically as competent an examiner using the questions as any expert in the trade, and more competent than a trade expert without the questions.

THE TECHNIQUE OF TRADE TEST CONSTRUCTION

ORAL TRADE TESTS

In order to formulate good oral trade test questions of the one-word-answer type one need but remember the simple rule: Decide first what is to be the answer to the question, and then, by administration of the question to tradesman, progressively revise the form of the question until there are no other allowable correct answers than the expected one (or synonomous or alternative answers).

The gain in objectiveness of administration and scoring made possible by the one-word-answer oral question may be shown by the two following examples of army oral trade tests on the same trade, issued almost a year apart. The first, or earlier one, has the advantage over ordinary tests used in industry of being compiled from questions which have been given a standardization tryout on workmen in the trade. The second, or later one, has all the additional advantages of almost perfect objectivity of scoring.

1. ARMY ORAL TRADE TEST—PATTERN-MAKER, WOOD

Issued April 17, 1918

1. Why is draft put on patterns?
   Ans. To enable the molder or foundryman to draw the pattern cleanly from the *mold* (sand). *Score* 4.

Digitized by Google

2. What is meant by rechucking a piece of wood in the lathe?
   Ans. (1) When a turned piece cannot be completed in one operation, it is removed from the lathe and set up again in such a position that other parts, or the reverse side of the piece can be turned. *Score* 4.
   (2) Reversing and turning the finished side toward the face plate. *Score* 4.
3. What is the least draft per foot in good molding practice?
   Ans. (1) One-sixteenth to one-eighth of an inch per foot. *Score* 4.
   (2) One-thirty-second of an inch per foot. *Score* 2.
4. What are two advantages of skeleton patterns?
   Ans. (1) *a.* Save labor.
   b. Save time.
   c. Save material. (Any two) *Score* 4
5. When gluing up stock for a pattern, which sides of the boards, as related to their position in the tree, should be placed together to keep the pattern from warping?
   Ans. The like sides (two heart sides) (two outsides). *Score* 4.
6. Explain how to glue the end grain of wood.
   Ans. (1) Clean the grain, size with hot glue, and when dry coat with hot glue; place the pieces together and clamp. *Score* 4.
   (2) Size it with glue and then glue it again. Rub together; glue and clamp. *Score* 4.
7. State two ways provided by the patternmakers for making cores.
   Ans. (1) *a.* By core boxes (solid boxes).
   b. By sweeps (strickles). (Both required). *Score* 4.
8. In making a pattern from which a brass working pattern is to be made, how much shrinkage should be allowed to insure the proper size of gray iron castings to be made from the brass pattern?
   Ans. (1) First allow 3/16 inch to the foot to take care of shrinkage of brass pattern, and then 1/8 inch to the foot to take care of shrinkage on gray iron castings off brass pattern; total of 5/16 inch allowance to the foot. *Score* 4.
   (2) 5/16 inch. *Score* 4.
   (3) 3/16 inch. *Score* 0.
9. How much shrinkage in the diameter would there be to an aluminum casting two feet in diameter and one-half inch thick?
   Ans. Three-eights to five-eighths of an inch. *Score* 4.
10. State how to find the mitering angle for any number of arms in a pattern for a wheel.
   Ans. (1) By dividing 360 (the number of degrees in a circle) by the number of arms. (NOTE: specific example is a sufficient answer). *Score* 4.
   (2) Divide the periphery or circumference of the pattern into as many equal parts as there are arms on the wheel. *Score* 4.
   (3) With trams (dividers). *Score* 2.

Digitized by Google

11. In making a solid cylinder pattern, on which side would you put the tapered core print?

Ans. (1) Cope side. *Score* 4.
(2) Drag side. *Score* 0.

RATING THE CANDIDATE

Rate N men scoring less than 15.
Rate A− men scoring 16 and 17.
Rate A men scoring 18 to 31.
Rate A+ men scoring 32 and 33.
Rate J− men scoring 34 and 35.
Rate J men scoring 36 to 39.
Rate J+ men scoring 40 and 41.
Rate E men scoring 42 or more.
There will be no E− or E+ rating.

2. U. S. EMPLOYMENT SERVICE TECHNICAL INTERVIEW. PATTERN-MAKER, WOOD

Issued February 2, 1919

1. What wood, besides pine, is most commonly used for making small patterns?
   Ans. Mahogany.
2. With what is the surface of a pattern coated to keep it from getting damp and warping?
   Ans. Shellac.
3. How is a pattern made so that it can be drawn out of the sand easily?
   Ans. Draft.
4. What is the box called in which the sand for a mold is rammed up?
   Ans. Flask.
5. For what is the allowance made on a pattern-maker's rule?
   Ans. Shrinkage (contraction).
6. What do you call the part of the pattern which is *above* the parting line on a two-part flask?
   Ans. Cope.
7. What tool do you use to lay out a pattern with a 3-foot radius?
   Ans. Trammels.
8. In making a large ring pattern, what are the separate parts called?
   Ans. Segments.
9. What do you call a mold that has a metal face to harden the casting?
   Ans. Chill.
10. What do you put on a pattern to support the core in the proper position?
    Ans. Prints.

EXAMINER: Ask at least six questions; more, if necessary, to accurately judge the candidate's ability. Score only the results of the six questions you ask first. Vary as much as possible your selection of the first six questions.

Digitized by Google

ASSEMBLY AND REVISION OF ORAL TRADE QUESTIONS AND PICTURE TRADE TEST QUESTIONS

Before assembling any questions from the experts in the field, one should first determine just what the trade is, in order to fit the questions to the trade. The best preparation for assembling questions on a trade is the acquirement of a large amount of knowledge concerning the trade.

Ordinarily from three to five experts in the trade, as defined below, should be employed in the assembly process. Psychologically the reason for this is simple; what will not be thought of by one expert will probably be thought of by another, so that ultimately, by the addition of but a few questions to the list by each expert, a lengthy list of good questions can be secured. The questions obtained from Expert No. 1 should be first submitted to Expert No. 2, requiring him to answer them in the same manner as if the questions were being asked in the standardization process. The answers of Expert No. 2 will in a large sense indicate whether the questions submitted by Expert No. 1 are ready for standardization. Expert No. 2 will obtain an adequate idea of the type of question desired from thus seeing what questions have already been submitted. If Expert No. 2 fails to answer any question submitted by Expert No. 1, this question should be carefully investigated with the aid of Expert No. 2 to discover whether it was a lack of information on the part of Expert No. 2 or a lack of clarity in the question which prevented his understanding what was meant. In the latter case, it may be necessary to reword the question many times.

Each question should thus progressively get better and better as submitted from one man to another in turn in the assembly process. If necessary, after the questions have been submitted to some two or three men with progressive revision by the examiner (preferably retaining the unstandardizable forms of questions) the questions should be taken back from the shop to the office and be typed. Then whenever successive experts fail to give the expected answer, the various alternative forms hitherto given can first be administered, finally ending by securing that man's contribution to another revised form. At any time that the exact expected answer is not immediately obtained from a superior workman, the difficulty should always be investigated. The assembler must remember, however, that even the best experts should fail on some questions, for it is only by hav-

Digitized by Google

ing some very difficult questions that the expert can be picked out.

Having secured, by progressive revision, an acceptable set of questions, the questions should be typed ready for standardization. Out of, say, one hundred questions secured from the first three experts, at least forty or fifty of the best questions should have stood the test of revision and be acceptable for standardization. The final selection of this set of fifty questions is made by a final administration of the questions to a "critical" examiner in the office, preferably a person who has had standardizing experience in various trades, and who has had some assembly experience on this particular trade. The answers to the entire set of questions, when this person attempts to guess all possible correct and incorrect, sensible and absurd answers to each question, combined with the joint judgment of the assembler and examiner as to whether an adequate sampling of the various aspects of the trade has been made and whether the questions are short enough to be quickly and easily administered, will be adequate enough for making the final selection of fifty questions for standardization. The critical examiner, upon being told the true expected answer, should attempt to guess synonymous or ambiguous alternative answers. If, then, the assembler's questions have been so worded as to pin the guesser to "I don't know," to the correct answer if known, or to an admittedly random guess, they are ready for typing for standardization. It is preferable to arrange the questions in their seeming order of difficulty, the easiest first. Even a very experienced assembler will often make very bad mistakes in his judgment of the difficulty of particular questions. However, a few easy questions, put at the beginning of the list, are desirable to set at ease the subjects used for standardization.

STANDARDIZATION

Standardization, according to any procedure, consists in administering the entire set of standardization questions, say fifty in number, to a group of tradesmen and non-tradesmen, of four classes of *known ability*, defined as follows:

NOVICE: A person totally without formal trade experience. Used as a check on examinations to make certain that the answers to questions cannot be guessed by an intelligent non-tradesman.

APPRENTICE: A beginner or learner in the trade, as determined by length of

Digitized by Google

experience in the trade, the kind of shop, and the foreman's opinion of a man. This does not necessarily coincide with the formal four or five years period of apprenticeship of the trade unions but recognizes the fact that apprentices of exceptional ability and two years experience may be more valuable workmen than some journeymen of less intelligence and a larger number of years of experience. In terms of experience alone, following the army practice, in no trade is more than five years experience required for a person of *average* ability to become a journeyman. The word "average" is to be noted here as recognizing the only point of view consistent with the facts of the distribution of human abilities. Concretely, this means that in answer to such questions as, "How long does it take to learn this trade?" one can consistently answer by stating how long *on the average* it takes a person of *average* ability to learn the trade. Statements of time required to learn a trade have no meaning unless one considers human ability as a variable quantity and unless the particular degree of ability to learn is specified as well.

JOURNEYMAN: A workman past the apprenticeship period, but not yet an expert. A practical criterion for journeyman is: not capable of working alone without directions or not capable of giving orders to others. A man with the requisite experience in years for an expert but not so classed by his foreman is to be considered as a journeyman.

EXPERT: A man past a definite stated minimum number of years of experience, varying with the trade considered, and *in addition* so classed by his foreman. In the army work, a foreman of a shop was nearly always classed as an expert. A man capable of acting as a foreman, or capable of working alone on difficult jobs without directions from others, was ordinarily also classed as an expert.

It can be readily seen that there are no sharp dividing lines between these trade classes in just the same way as there are no sharp dividing lines between the normal and backward, or backward and feeble-minded in intelligence. The expert is but a higher degree journeyman; the journeyman but a higher degree apprentice. Even the line between tradesman and non-tradesman is not so sharp as it appears at first glance; a tradesman of but three days' experience is but a novice for nearly all trade test purposes; and many persons may possibly be found, such as vocational guidance experts, mechanical engineers, and the like, who have a smattering of knowledge in many trades. There is no difference of "stuff" out of which these different classes of men are made. These classes are but arbitrary classes which are to be fairly well recognized in some trades, at least, as forming the promotional steps of a workman whenever these promotional steps are based upon trade ability and not upon mere number of years of experience. In some trades the expert may have a radically different type of work to do, at least a part of the time, than the journeyman; while in others the apprentice may have

Digitized by Google

a radically different type of work from the journeyman. The apprentice bricklayer, so far as trade tests are concerned, is the hod carrier; the apprentice riveter, for the same purposes, is the "heater up." These are artificial distinctions and may not be agreed to by the workmen in the trade. Where they occur, they are very valuable for distinguishing the various classes of men. In many cases it is difficult to decide in which status to put a given man; in such cases all possible evidence should be taken into account and the men classified according to the best judgment thus resulting. Evidently, to be scientific, this *rating of men must be complete before any tests are given*. The best practice, moreover, if it is possible thus to secure sufficient subjects for standardization, is to select only men who are cleancut examples of journeymen, apprentices and experts.

Next we should consider the number of men of each such class to whom the questions should be administered in standardization. In the army, an effort was made at first to administer the standardization questions to twenty each of apprentices, journeymen and experts and to forty novices. Later experience showed that even with the total resources of all workers in a great manufacturing district such as Newark or Pittsburgh to draw from, often not over a half dozen apprentices were to be found. The requirements in numbers were therefore changed as follows: 40 novices, 10 apprentices, 30 journeymen, and 20 experts. These proportions conform more nearly to the proportions in which the various classes are to be found in industry. Not so large a proportion of experts are generally to be found in any one shop. In a single shop, making up tests for the employment office of that shop, it will often be necessary to test all men in the shop if time permits. If there are, say, two rooms and twice as many men as it seems desirable to test, the best results will be obtained if an impartial sampling is taken of the men in both rooms, such a sampling as is to be easily obtained by testing all men in Room 1 whose last names begin with A to J inclusive, and in Room 2 all men whose last names begin with K to Z inclusive.

The order in which the various groups are tested is worth considering. Preferably a few apprentices are to be tested first. A low percentage of "passes" insures that the test is *difficult* enough to differentiate between different grades of ability. These should be followed by tests of a few experts. Ordinarily there should be no overlapping in total crude scores between these two groups. If overlapping

Digitized by Google

appears, the test must be critically examined, since it is then likely to be unsatisfactory.

Following the administration of the questions to the men of different classes of ability the questions are each statistically evaluated, only those being retained for the final examination set which fulfill the requirements of (1) statistically differentiating between the different degrees of trade ability, a graded series of difficulties being used; (2) being administratively acceptable as questions which will readily be answered as expected by a majority of expert tradesmen.

PICTURE TRADE TESTS

Picture trade tests are constructed in almost the same way. Since they are but oral tests, making use of pictures in place of descriptive phrases or definitions, all the points previously mentioned in respect to oral trade tests apply to the picture tests.

Care should be taken so to word the questions that the exact point of a picture referred to by the question shall be immediately recognized by the subject without requiring the examiner to point to any part of the picture. Care must therefore be taken that the pictures used for illustration be clearcut pictures. It is often possible to secure pictures, clipped from trade catalogs, which will clearly illustrate certain parts of machines and not others; in these cases, the questions should refer only to the clear part of the picture, other pictures being secured to illustrate the other parts of the machine. Of particular value are the pictures often found in catalogs, used to illustrate special appliances of machines, in which the background of the machine is purposely faded out in the photographic process to make the desired part of the machine stand out clear and distinct.

There is little advantage in having photographs specially made for a picture test. Ordinarily a very great abundance of good pictures of machines and parts may be found in trade catalogs. White ink, containing china white, is to be used for blotting out any undesirable lines, lettering or unessential backgrounds on the catalog pictures. The pictures used for the final series should not, in general, show any more details of the machinery than is necessary to illustrate the principles to which the questions refer. For instance, if one desires merely the name of a tool, one should select a picture which gives merely the outside outlines of the tool. In preparing catalog pictures for being photographed, care must also be taken to see that

Digitized by Google

photographs which have been originally photographed and developed "light" or "dark" are arranged together if best results are to be obtained.

The assembler of picture trade tests must usually know more of the trade than is required of the assembler of oral tests, since the former is required to judge largely for himself which pictures are best to use. This applies also to performance tests, wherein the assembler must be reasonably competent to judge whether a projected task involves enough of the essential tools and operations of a trade to be a fair sampling of the tradesman's skill and knowledge.

Once assembled and revised in a manner identical with that used in revision of an oral trade test, the picture test is ready for standardization. If both forms of the test are being constructed, much time may be saved, and a comparison of the respective reliabilities of the two forms of test may be secured, by standardizing both forms of the test on the same subjects at the same standardization examination.

Picture tests are not ordinarily made with the expectation that they will be better or quicker methods of examination than the oral method. Rather is their value to be sought in overcoming language difficulty or in checking up on suspected cases of coaching. To secure the same amount of information in regard to literate applicants' trade abilities, the picture test usually requires more time than the oral test. The generally safe rule to follow, when in doubt whether to use a picture or oral test, is to use both.

PERFORMANCE TESTS

Performance tests are more like picture tests than like oral tests. The elements for a performance test are chosen in much the same way as elements, *suitable for reproduction by pictures*, are chosen for a picture test, the only difference being that such tasks nust be chosen as will lend themselves to *easily and objectively measurable* scoring points. For measurement of the finished products a *series* of "limit" gages each varying from the preceding by thousandths or half-thousandths is preferable to the use of micrometers with their great possibilities for subjective errors of the scorer. Statistically, the ordinary limit gages of "dimension ± .001" are as poor a measure of a product for test purposes as dividing all humanity into short, tall and medium people. Thus partial credits are allowable, logical and necessary for best results in performance tests, which is not true in

Digitized by Google

case of the oral or picture tests where the burden of partial credits should be thrown upon the assembler who is to be compelled to assemble clear, understandable questions.

In the progressive assembly of performance tests, it is desirable to make use of both apprentices and experts, in order to determine what scoring points will yield differentiation of trade classes. Any scoring point which yields much overlapping of these classes is likely to be of little value, although more leeway may be allowed here than in the case of either of the other two forms of test. The product must be changed and rechanged during revision until a scorable and diagnostic product is discovered. Since a product is limited in its number of dimensions, it will be seen how difficult it is to devise a good performance test. Scoring points in regard to tools used, manner of holding tools, and the like,—deviating from the desirable method of scoring only on the finished product, which method does not require the services of an observer while the product is being produced,—must often be resorted to in this form of test, in order to get enough scoring points to differentiate adequately between different classes of workmen. The time required for the *total* operation must always be carefully taken and preserved, quality and quantity of production being the two generalized measures of the workman's skill in fashioning his product. The norms are always to be given in terms of quality X produced in time Y, or some composite index involving both of these variables. This is not necessary in the case of either oral or picture tests as the tests are given without a time limit, emphasis being placed upon the workman's "possession or non-possession" of the "essential" elements of trade knowledge and skill. Since the oral and picture tests have not been given with a time-limit, it is possible that with them also, better results might be obtained by reading the questions at a definite, predetermined rate and requiring the applicant to write his answers within a given number of seconds.

Copious notes, in addition to a minimum number of standard recorded observations of the standardization on workmen, should be taken in the standardization of performance tests. The best performance test is one in which mere measurements of the product will yield the desired differentiation; but if the final statistical evaluation should show that the workman's selection of tools, manner of handling tools, etc., are required as scoring points, these will be valueless unless carefully taken for every workman used in standardiza-

Digitized by Google

tion. For the assembly process, twice as much time should be used for the performance tests as for the oral or picture tests. Several times as long will of necessity be taken for both standardization and scoring, not to mention a proportionately longer time in the administration. In all trades save a very few which are almost wholly matters of skill, such as that of automobile driver and stenographer, performance tests have yielded *per unit of time spent in examination* or original assembly and standardization but a fraction of the examination value yielded by oral or picture tests. They have, from every viewpoint save that of dispelling an obstinate foreman's prejudice that "a man should be tested by giving him a job to do," but little claim to merit as tests for most tradesmen.

THE WRITTEN TRADE TEST

In the standardization process the administration of the questions to novices by means of written examinations proved highly successful in the army procedure. This is due, no doubt, largely to the fact that, for practical purposes, novices may be categorically classified as "knowing" or "not knowing" the items of information called for in the oral questions. If unable to answer them, they were seldom willing to make guesses. On the other hand, it has been found that tradesmen are very likely to question the correctness of some of the questions used, and it seems doubtful whether the method would have so high a degree of success with tradesmen as with non-tradesmen. Nevertheless, it is probably the case that much of the prejudice of the army camp examiners against a written method was the result of the reflection of the point of view of the practical man in industry, namely that "a man may be able to tell about his trade, but not be able to write about it." One can readily see that this is but an extension of the belief of such persons that a workman may be able to do a job, but not to talk about it. Statistics would probably show that there is a high correlation not only between trade skill and ability to talk about the trade, if the proper questions are used, but also between trade skill and ability to write about the trade, if, again, the proper questions are used.

The chief value in using any written trade test method, is, of course, the saving that results both in the actual examination time and in the time required to score the subject's replies. One examiner can examine five hundred men by written methods almost as quickly

Digitized by Google

as a single man by oral methods. The scoring may be done with great rapidity by stencil and by clerical help. Again, if the questions have been adequately standardized before administration, a definite, predetermined set of answers eliminates all subjective judgment on the part of the scorer.

One further advantage, only slightly used in school examinations, is the possibility of examining large numbers of persons at one time, allowing all who are evidently well above the minimum basis, to pass without further question and reserving those who make low marks for further detailed individual examination by oral or other methods. Of all the above methods of trade examination, the one-word-answer form of test is probably most readily adaptable to vocational school use. The only material needed for their assembly is paper and pencil. The questions, either printed or mimeographed with blank spaces following each question for the answer, are almost self-administrative. The scoring is thoroughly objective and rapidly done by means of stencil. This form of test is being given a thorough trial in the army E. and R. schools, both in tests of vocational aptitude and as final examinations, or tests of proficiency.

Another form of written test, easily adaptable to vocational school use because of its extreme ease of construction, is the multiple choice test. This test has not yet been given sufficient trial to have its merits actually determined. Among its obvious advantages are:

(1) The elimination of all writing. (2) The administration of more questions per unit of examination time. (3) Perfect objectivity in scoring, which is done by stencil. (4) Less scoring time per question than by the recall or picture method. (5) The possibility of using questions not adaptable to the oral or picture method, of discriminating between good and bad practice, and of making judgments as to the more preferable of two or more alternative methods, and the like.

Some light on the relative merits of these various methods will be provided by the results of the following three experiments.

THE BRICKLAYERS' TEST

An opportunity was given, through the courtesy of the Pittsburgh Bricklayers' Union, to experiment with a written bricklayers' examination devised on the well-known multiple choice method.[1]

[1] Chapman, J. C. and Toops, H. A. "A written Trade Test: Multiple Choice Method," *Journal of Applied Psychology*, Vol. 3, No. 4, 1919, pp. 358–365.

Digitized by Google

The men were all examined in one large assembly hall. There was therefore no possibility of arriving at incontestable measures of each man's trade status. Four questions requiring definite answers in regard to each man's trade experience and trade leadership ability were typed at the top of each test blank, and were answered by all. These answers were subjectively evaluated independently by two examiners and the status thereby determined. It is not to be expected that this method would result in a status comparable to the status given a man by his foreman, since there is present the element of a man's personal judgment of himself, and this is known to be highly unreliable. Various methods of selection of questions were tried out. It is unnecessary to give here the detailed explanations of the procedure. The following principles seem to be justified as a result of this experiment:

1. A few extremely bad questions will not appreciably change the relative marks of men if a large number of questions is used.

2. To secure the same differentiating ability this method of testing would seem to require a larger number of questions than with the corresponding oral form of test.

3. The eighteen best questions, chosen as conforming to the ideal type of graph distribution of correct answers,[1] are considerably better in differentiating the known groups into their proper classes than the entire set of sixty-two unselected questions.

4. Other things being equal, the larger the number of difficult questions in a written test, the better the test. Statistically this amounts merely to a restatement of the principle that the greatest reliability of individual scores is obtained when fifty per cent of a given test group pass and fifty per cent fail a given question.

5. Too many easy questions, as also too small a total number of questions, tend to cause (*a*) novices and apprentices to occupy too much space in the total scoring range and (*b*) journeymen and experts to "pile up" in the upper limits of the scoring points. It follows from the principles above enumerated that a large number of easy questions may, under certain conditions, be more generally efficient than a very much smaller number of difficult questions.

6. The results from the novices, persons so chosen that they should have been able to answer many of the questions if such ques-

[1] "The Making of a Trade Test," *Personnel*, Vol. 1, No. 7, 1918.

Digitized by Google

tions are capable of being guessed easily, show that this type of examination cannot be "bluffed."

7. The tradesman are, however, unwilling to make random guesses in this type of examination if they are uncertain of the correct answer. For this reason it seems that little is to be gained by scoring the test by any arbitrary scoring formula such as $S=R-W$. Especially when given with a work limit, as was this test is the discount for the guessing privilege unnecessary, since the relative ranks of the men remain the same, all questions being presumably attempted by all men. If, however, the test were to be given as a time limit test, a scoring formula might always yield better results than by counting as the score the Rights alone. The best theoretical discount, of course, is that discount which would be determined by solving for C in the scoring formula $S=R-C\cdot W$, in which C is a constant, the ratio of the partial regression coefficient of W, divided by the partial regression coefficient of R, with S as an independent criterion of the measure of trade ability of the men tested.

THE FARMERS' TEST

To further test the multiple-choice plan of written examination, a farmers' test was similarly prepared. The one hundred and thirteen questions evaluated in standardization are shown below.

This test represents the one hundred and thirteen questions left for evaluation after seventeen of the original questions had been at once discarded as a result of the administration of the questions to farmers, and later two more, Nos. 84 and 85, of the list of one hundred and fifteen originally retained.

The numbers in the right-hand margin indicate the number of the four choices which is the correct answer as judged by Ohio farmers.

FARMERS' TEST

Each of the numbered questions below can be made to read correctly when one of the four words in the parentheses is underscored. Underscore the one proper word which will make the sentence read correctly. *Do not underscore more than one word.* If you are not sure which word is right, make the best guess you can; you will be discounted if you do not answer every question. You may have all the time you wish but accuracy is highly important. The two samples are marked correctly.

Samples
A. The end-gate is part of a (barn, fence, wagon, buggy).
B. The Berkshire is a breed of (sheep, horses, cattle, hogs).

Digitized by Google

INFORMATION QUESTIONS:

1. The Babcock Test is used to determine (the health of an animal, the condition of the soil, strength of a solution, butter-fat of milk) 4
2. A hame is part of a (binder, tractor, harness, potato-digger) 3
3. A scythe is usually driven by (steam-power, electric-power, horse-power, hand-power) 4
4. A hay-tedder is used in (loading, drying, baling, unloading) hay 2
5. The Guernsey is a breed of (turkey, sheep, cattle, horse) 3
6. Hens usually set on (6, 10, 15, 20) eggs 3
7. A cow gives milk from (2, 4, 6, 12) teats 2
8. In driving a team, the command, "Haw," means to (turn to the left, turn to the right, go straight ahead, back up) 1
9. Acid phosphate is a (drug, food, drink, fertilizer) 4
10. Ensilage is a (feed, flower, tool, disease) 1
11. A whiffletree is a (plant, tool, animal, part of a wagon) 4
12. Roup is (part of a harness, a garden vegetable, disease of chickens, a stock food) 3
13. Gaps is a disease of (poultry, sheep, cattle, swine) 1
14. The first crop planted in spring is (corn, oats, beans, buckwheat) . . . 2
15. An average yield of wheat per acre is (10, 18, 30, 50) bushels 2
16. A capon is a (chicken, fruit, garden-vegetable, field-crop) 1
17. Corn is planted in (March, May, June, July) 2
18. Corn stover is used in feeding (chickens, cattle, hogs, poultry) 2
19. Casein comes from (meat, milk, corn, fruit) 2
20. The deepest-rooted forage plant is (timothy, clover, alfalfa, millet) . . 3
21. The most hardy fruit is the (apple, peach, plum, cherry) 1
22. Chess, or cheat, injures (oats, barley, rye, wheat) 4
23. A horse is usually broken to work at age of (2, 3, 4, 5) years 2
24. A good yield of hay is (1, 2, 3, 4) tons per acre 2
25. The longest strawed grain is (wheat, barley, rye, oats) 3
26. A good hand will husk (25, 50, 100, 150) bushels of corn in a day . . 2
27. The heaviest milking breed of cows is (Jersey, Hereford, Shorthorn, Holstein) 4
28. Animals that chew the cud are (horses, hogs, cattle, dogs) 3
29. A snath is the handle of a (plow, rake, fork, scythe) 4
30. Legumes are a (family of plants, part of a harness, fixtures in a stable, attachment to a corn-shredder) 1
31. A man and team should be able to care for and cultivate (10, 40, 60, 80) acres of corn in a season. 2
32. One breed of hogs is the (Hereford, Minorca, Merino, Duroc) 4
33. Wheat is cut with a (mowing-machine, scythe, binder, rolling-cutter) . 3
34. A common make of binder is (Iver-Johnson, McCormick, Ford, Racine) 2
35. A tool used in cultivating a garden is (sulky, drill, hoe, roller) 3
36. The kind of horse most used on a farm is (road, coach, draft, race) . . 3
37. A good cow feed is (straw, alfalfa, timothy-hay, oats) 2
38. A good hog feed is (corn, fodder, sorghum, peas) 1

Digitized by Google

39. A fowl used on most farms is (chicken, goose, turkey, pheasant) . . . 1
40. A tool used in hay-making is (binder, mower, scythe, harrow) 2
41. Field corn ripens in (June, July, September, November) 3
42. Winter wheat ripens in (March, April, July, October) 3
43. In preparing ground for seeding, a farmer uses (harrow, seed drill, seeder, subsoiler) 1
44. Hay should be cured (15-minutes, 30-minutes, 1-day, 1-week) before being put in the mow. 3
45. A rolling coulter is part of a (harrow, binder, rake, plow) 4
46. A share is part of a (binder, harrow, plow, tedder) 3
47. A doubletree is part of a (harness, wagon, sprayer, posthole-digger) . . 2
48. Ditching can be done most easily in (spring, summer, autumn, winter) 1
49. Cows are usually milked (2, 3, 4, 5) times a day 1
50. A fat market hog will weigh about (50, 100, 250, 500) pounds 3
51. An average feeding steer will weigh (300, 600, 900, 1200) pounds . . . 3
52. A moldboard is part of a (scythe, engine, plow, corn-cutter) 3
53. A swath is made by a (harrow, plow, binder, mowing-machine) . . . 4
54. A pitman rod is part of a (mowing-machine, corn-planter, roller, doubletree) 1
55. A disk is part of a (roller, roll-cutter, binder, mowing-machine) . . . 2
56. A reel is part of a (mowing-machine, binder, roller, disk-cutter) . . . 2
57. A drag (has shafts, has a tongue, is tongueless, has a pole) 3
58. A 6-foot binder is pulled by (2, 3, 5, 7) horses 2
59. Tomato plants should be planted about (1, 2, 4, 8) feet apart in the row 3
60. Vinegar is made from (grapes, blackberries, raspberries, apples) . . . 4
61. Hog house doors should be at least (1, 1¼, 1½, 2½) feet high 4
62. Sheep are best raised on (wet, level, hilly, weedy) ground 3
63. A field from which the wheat has been cut is called a (wheat-field, pasture, meadow, stubble-field) 4
64. There are (40, 60, 100, 160) square rods in an acre 4
65. A jobber is sometimes used to plant (wheat, corn, oats, rye) 2
66. A clevis is part of a (scythe, doubletree, jobber, clover-seeder) 2
67. A kingbolt is part of a (wagon, automobile, harrow, buggy). 1
68. A plant grown both for pies and for cow-feed is (rhubarb, peas, pumpkins, turnips) 3
69. Post holes should be dug (1, 3, 5, 7) feet deep 2
70. It requires (1, 2, 3, 4) bushels of oats to seed an acre 2
71. The first crop harvested in a season is (timothy, oats, clover, winter-wheat) 3
72. Pigs are usually weaned at the age of (2, 6, 10, 15) weeks 3
73. Pigs at birth weigh about (1, 3, 4, 5) pounds 2
74. The incubation period for chicken's eggs is (2, 3, 4, 6) weeks 2
75. A team plows (2, 3, 4, 5) acres of ground per day with a 14-inch walking breaking plow 1
76. A week-old calf should be fed (2, 4, 6, 10) quarts of milk at a feed . . 2
77. Thumps is (part of a wagon, a garden implement, a symptom of disease in pigs, an unsoundness in the horse) 3

Digitized by Google

78. The most valuable forage for dairy cows is (clover, timothy, oat-hay, alfalfa) 4
79. A threshing outfit will thresh (500, 1000, 2500, 4000) bushels of wheat in a day 2
80. Weeds most seriously reduce the yield of (corn, oats, wheat, hay) . . . 1
81. A manure spreader covers a strip of ground (3, 6, 10, 12) feet wide . . 2
82. An ordinary smoothing harrow is (6, 10, 15, 18) feet wide 2
83. Hilled corn should be planted (6 in., 8 in., 3 ft., 6 ft.) apart in the rows . 3
86. Plowing the soil for corn should be done to a depth of (2, 6, 10, 18) inches 2
87. To conserve the moisture in the soil a (drill, drag, harrow, cultivator) should be used immediately after plowing 2
88. A 1500-pound horse will eat (5, 10, 20, 50) pounds of hay a day . . . 3
89. Corn should be cultivated to a depth of (1, 3, 7, 10) inches 2
90. A mowing machine cuts a path about (1, 3, 5, 8 feet wide) 3
91. A common width of furrow is (6, 8, 10, 14) inches 4
92. Red clover ordinarily yields (1, 2, 3, 4) crops per season 2
93. The best laying hens will lay (100, 200, 300, 400) eggs per year 2
94. To produce good crops, wet soil should be (tiled, subsoiled, irrigated, dynamited) 2
95. A horse at work is fed (4, 6, 8, 10) quarts of oats at a feed 1
96. Four-inch drain tile are usually (8, 12, 18, 24) inches long 2
97. It requires (3, 5, 7, 9) acres of grass to furnish pasturage for a cow during the pasturage season 1
98. One of the birds that hops over the ground is (robin, blackbird, crow, quail) 1
99. It takes about (1, 4, 8, 12) pounds of binder twine to bind an acre of good wheat 2
100. The tree of these four that shows its leaf earliest in the spring is (oak, ash, maple, hickory) 3
101. A veal calf should be marketed at the age of (2, 4, 6, 10) weeks . . . 3
102. The first fruit tree of these to blossom in the spring is (cherry, apple, pear, quince) 1
103. Corn is planted to a depth of (1, 3, 6, 8) inches 2
104. A fattening steer will gain (½, 1, 3, 5) pounds a day in weight 3
105. Three horses are hitched together by means of an (evener, pole, single-tree, doubletree) 1
106. It takes (200, 500, 1000, 2000) pounds of fertilizer per acre of wheat . 1
107. It requires (4, 8, 14, 20) bushels of potatoes to plant an acre 2
108. A week old calf is fed (once, twice, 3-times, 4-times) a day 2
109. A peach tree should begin to bear at age of (1, 3, 5, 7) years 2
110. It takes (1, 2, 3, 4) bushels of wheat to sow an acre of good ground . . 2
111. Apple trees should be planted (10, 20, 30, 50) feet apart 3
112. Cabbage plants should be planted (4, 6, 10, 16) inches apart in the rows 4
113. In drilled wheat the rows are about (4, 7, 9, 12) inches apart 2
114. A farm horse should be shod in intervals of (2-weeks, 6-weeks, 3-months, 6-months) 3
115. To destroy lice, hogs may be dipped in a solution of (molasses, turpentine, lard, carbolic-acid) 4

Digitized by Google

The questions were administered to ten farmers living on Ohio farms and regarded as "successful" farmers; to twenty-one women novices living in the city, three of whom were born on farms; to twenty-four men novices, some of whom were college graduates and had lived for a long period of time on the farm, from five to eighteen years in their boyhood; to nine advanced students in a state college of agriculture, preparing to become high-school teachers of agriculture; and to thirteen high-school teachers of agriculture in the same state as the college of agriculture students.

Taking the one hundred and thirteen questions, without further selection, and setting up the usual critical scoring lines (43) between novices and "semi-skilled" farmers; between "semi-skilled" farmers and "successful farmers," all women novices, save one with previous experience of eighteen years on a farm, are rated as novices. None of the men novices received ratings above "semi-skilled." Of the nine advanced students of the state college of agriculture, eight are rated by the test as "semi-skilled" and one as a novice; of the thirteen high-school teachers of agriculture, ten are rated as "semi-skilled" and three as "successful farmers." Of the "successful farmers" all save one boy, included in the group, who was at the time attending high-school, are rated as "successful farmers." The results thus show a remarkably good differentiation of farming ability as far as it can be judged from the measure, which is not, of course, fully adequate.

The technical training of the college boys, subjected in all due fairness, it must be stated, to a somewhat different type of farming from that prevailing on the general farms of Ohio, was not sufficient to enable any of them to obtain ratings as "successful farmers." That the difference is not due wholly to the difference in the section of the country is proved by the fact that the high-school teachers of agriculture from the same state, and somewhat experienced in the actual farming operations of their vicinities, were able to make a higher average score, and one quarter of them to receive ratings as "successful farmers," judged by the Ohio standard. We may infer from this that the questions selected were "practical questions" in the sense that one must have had actual farm experience in order to make high scores on the test.

There are seventy questions of the set which no single one of the farmers missed. To the average man of the street such questions would be the ideal type of question. The reasoning would be some-

Digitized by Google

what as follows: If no single farmer missed any one of the seventy questions, then failure to pass any considerable number of the questions would undoubtedly prove a serious lack of farming skill or knowledge. Philosophically, such owes its origin to the type of thinking characteristic of the trick question so popular with many foremen. Individual scores were computed on the seventy easiest questions—easiest for Ohio farmers—and the results graphed as usual. With the critical scoring points set up, farmers were separated from men novices, although there was little leeway in scoring points between the two groups. Inexperienced men novices and the experienced novices were here grouped together indiscriminately in scores. Thus, the set is not a bad examination, but could be very much improved by proper selection of questions. Undoubtedly it is only by having a very large number of questions that the rather good differentiation has been obtained. It must be remembered that even "trick questions" will differentiate the various trade classes if we use enough of them. The difficulty in industry with the "trick question" is that the foreman is prone to hire or not hire a man on the basis of his answers to one, or at most, two or three of these "trick questions."

After the seventy easiest questions had been evaluated as above, the forty-three most difficult questions were similarly evaluated. It is astonishing to discover that the differentiation of groups secured by using the difficult questions is much inferior to that obtained from the easier ones. Men novices in general make scarcely any better scores than women novices; there is some overlapping of men novices with farmers; while four of the women novices made higher scores than the lowest farmer. This lack of differentiation is probably due to the small number of questions involved rather than to failure of difficult questions to secure the differentiation which difficult questions are usually expected to secure. The method gave very poor differentiation in spite of the fact that the questions were difficult. Examining our measure of the percentage of the total group who failed, it is a question whether the lack of differentiation with difficult questions is not perhaps due to the fact that the questions, if graphed, would present unfavorable graphs. Without graphing, but merely from inspection of a chart on which the number of failures by known classes was tabulated, a set of forty questions of good graph form was selected. Counting all the men novices as one

Digitized by Google

class and as superior to women novices in regard to their effectiveness as farmers, since but few of the men were totally inexperienced, the forty questions yield an excellent differentiation of the three classes, and a perfect separation of "persons living on a farm" from "persons not living on a farm." The results of this experiment confirmed the conclusions obtained in the bricklayers' written trade test, namely that (1) either a very large number of written questions must be used, or else (2) a careful selection on the basis of graphed questions must be made, and (3) when in doubt as to the choice between two questions seemingly alike in other respects, it is preferable generally to select the more difficult question.

A COMPARISON OF THREE FORMS OF WRITTEN TEST

In order to determine the relative merits of three types of examination method, the reactions of subjects to a test of general information were studied. There are two statistical aspects of the worth of any test: (1) its correlation with an independent criterion of the ability tested, and (2) its reliability, as measured by the correlation coefficient of reliability. Only the second of these two aspects has been investigated in this experiment.

Fifty items of information, designed to sample a very wide range of general information ability, were assembled and arranged in three forms of test: (1) recall test, or army oral trade test, with the one-word-answer form of question, (2) recognition test, with the questions worded exactly the same as in the recall, but with five answers given to each question, only one of which was correct, (3) true-false test, the material being arranged to make use as far as possible of no new concepts. The tests are shown on the following pages.

GENERAL INFORMATION TEST. RECALL

Each question below can be answered with a single word or a number. Write the answer in the blank space following each question.

Samples
Who was the first president of the United States?
Ans. Washington.
What color are ripe strawberries?
Ans. Red.

1. What letter designates the note on the bottom line of the staff in music?
   Ans. E.
2. What is the smallest fraction of a cent in which stocks are quoted?
   Ans. ⅛.

Digitized by Google

3. How many minutes does it take light to travel from the sun to the earth?
   Ans. 8.
4. By what sculptor was "The Thinker" made?
   Ans. Rodin.
5. What was the nationality of William of Orange?
   Ans. Dutch.
6. Who, in American history, is known as "Old Hickory"?
   Ans. Jackson.
7. How many feet does an object fall from a position of rest in the first second?
   Ans. 16.
8. How many senators represent each state at Washington?
   Ans. 2.
9. What was the nationality of Rembrandt?
   Ans. Dutch.
10. What is width in inches of a hand by which the height of horses is measured?
    Ans. 4.
11. By whom was Jeffries defeated in the prize fight at Reno?
    Ans. Johnson.
12. Who founded the Mormon religion?
    Ans. Smith.
13. What wood was found best for aeroplane construction?
    Ans. Spruce.
14. Who said "I came, I saw, I conquered?"
    Ans. Cæsar.
15. How many teams are there in the National Baseball League?
    Ans. 8.
16. Who is the father of modern socialism?
    Ans. Marx.
17. In what city was President McKinley assassinated?
    Ans. Buffalo.
18. What is the key of music which has no sharps or flats?
    Ans. C.
19. What American admiral first visited Japan?
    Ans. Perry.
20. In what country has the story of William Tell its origin?
    Ans. Switzerland.
21. What is the largest city in South America?
    Ans. Buenos-Aires.
22. How many feet are there in a fathom?
    Ans. 6.
23. What is the smallest planet in the solar system?
    Ans. Mercury.
24. What acid is contained in vinegar?
    Ans. Acetic.
25. For what American poet was the armchair made from the spreading chestnut tree?
    Ans. Longfellow.

Digitized by Google

26. In what river are the Thousand Isles?
Ans. St. Lawrence.
27. By what insect is the yellow fever germ carried?
Ans. Mosquito.
28. On what mountain were the commandments received by Moses?
Ans. Sinai.
29. What stick is generally used in golf when on the green?
Ans. Putter.
30. From what animal is ambergris secured?
Ans. Whale.
31. How many legs does an insect usually have?
Ans. 6.
32. Who created the character of Ichabod Crane?
Ans. Irving.
33. By whom was the "Angelus" painted?
Ans. Millet.
34. What breed of dairy cow gives the greatest amount of milk?
Ans. Holstein.
35. What chemical element is the basis of graphite?
Ans. Carbon.
36. What is the name of the human bone which extends from the knee to the hip?
Ans. Femur.
37. What Biblical character said, "Your people shall be my people?"
Ans. Ruth.
38. Who discovered the Mississippi River?
Ans. De Soto.
39. What mythological beauty was the cause of the Trojan War?
Ans. Helen.
40. During what war did the charge of the light brigade take place?
Ans. Crimean.
41. In what country is most of the world's sisal crop grown?
Ans. Mexico.
42. On how many hills was Rome built?
Ans. 7.
43. What wrestler is famous for the "scissors hold"?
Ans. Stecher.
44. In what play does the character of Portia appear?
Ans. The Merchant of Venice.
45. What instrument is used to test the strength of a storage battery solution?
Ans. Hydrometer.
46. What state is called the "mother of presidents"?
Ans. Virginia.
47. What breed of hen is best for egg production?
Ans. Leghorn.
48. To what country does Greenland belong?
Ans. Denmark.

Digitized by Google

49. What state contains a section known as the "Panhandle"?
Ans. Texas.
50. In what city is the Smithsonian Institute?
Ans. Washington.

GENERAL INFORMATION TEST. RECOGNITION

Read each question carefully, and then underscore the *one* correct answer.

Samples
How many inches are there in a foot?
Ans. (8, 10, 12, 14, 16).
Who was the first president of the United States?
Ans. (Lincoln, Adams, Franklin, Grant, Washington).

1. What letter designates the note on the bottom line of the staff in music?
Ans. (A, E, G, B, C).
2. What is the smallest fraction of a cent in which stocks are quoted?
Ans. (1/2, 1/3, 1/8, 1/5, 1/10).
3. How many minutes does it take light to travel from the sun to the earth?
Ans. (4, 8, 12, 16, 20).
4. By what sculptor was "The Thinker" made?
Ans. (Michael-Angelo, Bartholdi, Borglum, St.-Gaudens, Rodin).
5. What was the nationality of William of Orange?
Ans. (Dutch, Scotch, French, English, Irish).
6. Who, in American History, is known as "Old Hickory"?
Ans. (Johnson, Jackson, Lee, Lincoln, Boone).
7. How many feet does an object fall from a position of rest in the first second?
Ans. (8, 12, 16, 20, 24).
8. How many senators represent each state at Washington?
Ans. (1, 2, 3, 4, 5).
9. What was the nationality of Rembrandt?
Ans. (German, Swiss, Belgian, Dutch, Austrian).
10. What is width in inches of a hand by which the height of horses is measured?
Ans. (4, 6, 8, 10, 12).
11. By whom was Jeffries defeated in the prize fight at Reno?
Ans. (Willard, Corbett, Fitzsimmons, Dempsey, Johnson).
12. Who founded the Mormon religion?
Ans. (Dowie, Young, Smoot, Smith, Penn).
13. What wood was found best for aeroplane construction?
Ans. (Hickory, pine, ash, spruce, hemlock).
14. Who said, "I came, I saw, I conquered"?
Ans. (Napoleon, Cæsar, Washington, Cromwell, Grant).
15. How many teams are there in the National Baseball League?
Ans. (4, 6, 8, 10, 12).
16. Who is the father of modern socialism?
Ans. (Trotsky, Debs, Penrose, Tolstoy, Marx).
17. In what city was President McKinley assassinated?
Ans. (Buffalo, Chicago, St. Louis, Rochester, Milwaukee).

Digitized by Google

18. What is the key of music which has no sharps or flats?
Ans. (A, C, D, E, G).
19. What American admiral first visited Japan?
Ans. (Farragut, Jones, Evans, Perry, Deucy).
20. In what country has the story of William Tell its origin?
Ans. (England, Switzerland, Spain, France, Italy).
21. What is the largest city in South America?
Ans. (Rio-de-Janerio, Lima, Santiago, Valpariso, Buenos-Aires).
22. How many feet are there in a fathom?
Ans. (4, 6, 8, 10, 12).
23. What is the smallest planet in the solar system?
Ans. (Venus, Saturn, Mercury, Neptune, Mars).
24. What acid is contained in vinegar?
Ans. (Acetic, citric, nitric, oxalic, carbolic).
25. For what American poet was the armchair made from the spreading chestnut tree?
Ans. (Whittier, Lowell, Tennyson, Longfellow, Bryant).
26. In what river are the Thousand Isles?
Ans. (Hudson, St. Lawrence, Delaware, Mississippi, Amazon).
27. By what insect is the yellow fever germ carried?
Ans. (House-fly, cootie, flea, mosquito, tick).
28. On what mountain were the commandments received by Moses?
Ans. (Ararat, Everest, Aetna, Mount-of-Olives, Sinai).
29. What stick is generally used in golf when on the green?
Ans. (Mashie, lofter, putter, brassie, cleak).
30. From what animal is ambergris secured?
Ans. (Elephant, whale, walrus, camel, shark).
31. How many legs does an insect usually have?
Ans. (4, 6, 8, 10, 12).
32. Who created the character of Ichabod Crane?
Ans. (Cooper, Dickens, Dumas, Irving, Stevenson).
33. By whom was the "Angelus" painted?
Ans. (Corot, Bonheur, Millet, DaVinci, Velasquez).
34. What breed of dairy cow gives the greatest amount of milk?
Ans. (Jersey, Guernsey, Ayreshire, Brown-Swiss, Holstein).
35. What chemical element is the basis of graphite?
Ans. (Carbon, lead, mercury, zinc, iron).
36. What is the name of the human bone which extends from the knee to the hip?
Ans. (Ulna, femur, humerus, cervical, clavicle).
37. What Biblical character said, "Your people shall be my people"?
Ans. (Esther, Rachael, Rebecca, Ruth, Hester).
38. Who discovered the Mississippi river?
Ans. (DeSoto, Balboa, Drake, Ponce-de-Leon, Cortez).
39. What mythological beauty was the cause of the Trojan War?
Ans. (Venus, Minerva, Helen, Ceres, Juno).

Digitized by Google

40. During what war did the charge of the light brigade take place?
Ans. (Napoleonic, War-of-the-Roses, Franco-Prussian, United States Civil, Crimean).
41. In what country is most of the world's sisal crop grown?
Ans. (India, Brazil, Mexico, China, Turkey).
42. On how many hills was Rome built?
Ans. (3, 5, 7, 9, 11).
43. What wrestler is famous for the "scissors hold"?
Ans. (Lewis, Stecher, Caddock, Roller, Demetral).
44. In what play does the character of Portia appear?
Ans. (King Lear, Twelfth Night, Julius Cæsar, The Merchant of Venice, Henry the Eighth).
45. What instrument is used to test the strength of a storage battery solution?
Ans. (Hydrometer, thermometer, barometer, altometer, speedometer).
46. What state is called the "mother of presidents"?
Ans. (Massachusetts, Maryland, Ohio, New York, Virginia).
47. What breed of hen is best for egg production?
Ans. (Rhode-Island-red, cochin, wyandotte, leghorn, barred-rock).
48. To what country does Greenland belong?
Ans. Norway, Sweden, Denmark, England, France).
49. What state contains a section known as the "Panhandle"?
Ans. (Arizona, Texas, Louisiana, Kentucky, California).
50. In what city is the Smithsonian Institute?
Ans. (Boston, Philadelphia, Pittsburgh, Baltimore, Washington).

GENERAL INFORMATION TEST. TRUE—FALSE

Examine each statement carefully. If the statement is true, underline the word, "TRUE"; but if the statement is false, underline instead the word, "FALSE."

Samples { There are sixteen inches in a foot . . . . . TRUE FALSE
Ripe strawberries are red . . . . . TRUE FALSE

1. The letter G is the note which is on the bottom line of the staff in music . . . . . TRUE FALSE
2. The smallest fraction of a cent in which stocks are quoted is 1/10 . . . . . TRUE FALSE
3. It takes light eight minutes to travel from the sun to the earth. . . . . TRUE FLASE
4. The statute of "The Thinker" was made by Bartholdi . . TRUE FALSE
5. The nationality of William of Orange is Dutch . . . . . TRUE FALSE
6. President Jackson is known as "Old Hickory." . . . . . TRUE FALSE
7. An object falls twenty feet from a position of rest in the first second . . . . . TRUE FALSE
8. Two senators represent each state at Washington . . . . TRUE FALSE
9. Rembrandt was of Dutch nationality . . . . . TRUE FALSE
10. A hand, by means of which the height of horses is measured, is six inches in width . . . . . TRUE FALSE

Digitized by Google

11. Johnson defeated Jeffries in the prize fight at Reno . . . TRUE FALSE
12. Dowie founded the Mormon religion . . . . . TRUE FALSE
13. Ash was found to be the best wood for æroplane construction . . . . . TRUE FALSE
14. It was Cæsar who said, "I came, I saw, I conquered." . . TRUE FALSE
15. There are ten teams in the National Baseball League . . TRUE FALSE
16. Marx is the father of modern socialism . . . . . TRUE FALSE
17. President McKinley was assassinated in Rochester . . . TRUE FALSE
18. The key of music which has no sharps or flats is C . . . TRUE FALSE
19. Perry was the first admiral to visit Japan . . . . . TRUE FALSE
20. The story of William Tell has its origin in Italy . . . . TRUE FALSE
21. The largest city in South America is Rio de Janeiro . . . TRUE FALSE
22. There are six feet in a fathom . . . . . TRUE FALSE
23. The smallest of the planets in the solar system is Venus . TRUE FALSE
24. Acetic is the acid which is contained in vinegar . . . . . TRUE FALSE
25. The armchair made from the spreading chestnut tree was made for Longfellow . . . . . TRUE FALSE
26. The Thousand Isles are located in the St. Lawrence River TRUE FALSE
27. The tick is the carrier of the yellow fever germ . . . . . TRUE FALSE
28. Moses received the commandments on the Mount of Olives TRUE FALSE
29. The putter is used in golf when on the green . . . . . TRUE FALSE
30. Ambergris is secured from the shark . . . . . TRUE FALSE
31. An insect usually has six legs . . . . . TRUE FALSE
32. Irving created the character of Ichabod Crane . . . . . TRUE FALSE
33. The "Angelus" was painted by Bonheur . . . . . TRUE FALSE
34. Guernsey cows give the greatest amount of milk . . . . TRUE FALSE
35. Carbon is the chemical basis of graphite . . . . . TRUE FALSE
36. The human bone which extends from the knee to the hip is called the femur . . . . . TRUE FALSE
37. Ruth is the Biblical character who said, "Your people shall be my people." . . . . . TRUE FALSE
38. Balboa discovered the Mississippi River . . . . . TRUE FALSE
39. Helen was the cause of the Trojan War . . . . . TRUE FALSE
30. The charge of the light brigade took place during the Franco-Prussian War . . . . . TRUE FALSE
41. Most of the world's sisal crop is grown in India . . . . TRUE FALSE
42. Rome was built on seven hills. . . . . TRUE FALSE
43. The wrestler, Demetral, is famous for the "scissors hold.". TRUE FALSE
44. Portia appears in the play, The Merchant of Venice . . TRUE FALSE
45. A barometer is used to test the strength of a storage battery solution . . . . . TRUE FALSE
46. Ohio is called the "mother of presidents.". . . . . TRUE FALSE
47. Leghorn hens are the best breed for egg production . . . TRUE FALSE
48. Greenland belongs to Iceland . . . . . TRUE FALSE
49. Texas contains a section known as the "Panhandle" . . . TRUE FALSE
50. The Smithsonian Institute is located in Washington . . . TRUE FALSE

Digitized by Google

The tests were given in all three forms to any one person at one sitting by the group examination method. Since results might conceivably be influenced by the order in which the forms were taken, all of the six possible combinations of orders were administered to different subjects of the same school classes by clipping together the test papers in the desired order of administration, hereinafter called "method." The order of the methods is as follows:

	No. of Cases
Method 1. Recall, recognition, true-false	39
Method 2. Recall, true-false, recognition	37
Method 3. Recognition, recall, true-false	14
Method 4. Recognition, true- false, recall	12
Method 5. True-false, recall, recognition	10
Method 6. True-false, recognition, recall	12
Total	124

The number of cases by methods is the same throughout, save in the calculation of average times taken on the tests where the frequencies are smaller because some subjects failed to record the time taken. The approximate times, taken to the last minute already recorded when a subject finished each test, were found by the simple expedient of the examiner marking on the blackboard the time elapsed at the end of every minute, the time of completion of each form being written on the test paper by the subjects. This procedure was not followed in two classes tested. The test was given by the work-limit method, accuracy and not speed being stressed in the directions. The subjects were summer school students at Teachers College. Since reliability is being investigated, sex differences or class differences need not be considered. The results are shown in Table I.

On all methods, the subjects made the lowest average score on the recall form of test, a higher average score on the recognition form, and a still higher average score on the true-false form. The true-false form is decidedly too easy a test for university students, there being from 76 to 81 per cent of passes, according to the method.

As shown by the average scores, the recall method profits by having one of the other forms given first. In methods 1 and 2, with recall given first, the average recall scores are 87 and 84 per cent of the average recognition scores; in methods 3 and 4, with recognition

Digitized by Google

TABLE I

AVERAGE SCORES, AVERAGE INTERCORRELATIONS OF EACH TEST RESPECTIVELY WITH THE OTHER TWO, RELIABILITY COEFFICIENTS OF HALVES OF TEST, AVERAGE TIMES IN MINUTES, BY METHODS AND BY TESTS

		Average Score on test			*Average Correlation of tests with other two tests in each case*			*Reliability Coefficient of the halves of tests*			*Average time in minutes*			
Method Number	*No. of Cases*	*Recall*	*Recognition*	*True-False*	*Recall*	*Recognition*	*True-False*	*Recall*	*Recognition*	*True-False*	*Recall*	*Recognition*	*True-False*	*No. of Cases on "time" only*
1	39	*29.1*	33.3	39.2	.872	.670	.811	.724	.487	.405	7.8	5.4	3.4	28
2	37	*30.3*	35.9	39.1	.818	.722	.868	.784	.668	.690	8.0	4.7	3.6	24
3	14	31.4	*32.4*	37.4	.938	.827	.852	.784	.785	.664	5.5	8.2	3.0	12
4	12	35.6	*36.9*	39.6	.926	.898	.924	.789	.530	.596	5.4	5.6	3.7	7
5	10	36.2	39.1	*40.3*	.358	.803	.432	.284	.271	.385	5.2	4.9	4.2	9
6	12	35.5	36.8	*38.2*	.968	.740	.762	.695	.675	.604	6.2	6.2	4.3	6
Total	124				*Average of six Average Intercorrelations*			*Total of 124 cases*			*Total Average Weighted Time*			86
					.813	.777	.775	.448	.385	.340	6.9	5.6	3.6	

given first, the average recall scores are respectively 97 and 96 per cent of the average recognition scores, and the average recall scores are larger than in the previous two methods; in methods 5 and 6, with true-false given first, the average recall scores are respectively 93 and 96 per cent of the average recognition scores, while the average recall scores are higher than in all but one of the four previously mentioned methods. The true-false test is subject to little or no improvability through previous tests, and the recognition a lesser amount than the recall form of test.

In the absence of a better criterion, we may say that that test is the best measure of general information, which, on the average, correlates highest with the other two tests. With this criterion, recall ranks first, recognition second, and true-false last.

Digitized by Google

The recall form, with fifty questions used in all three forms, will yield the highest reliability coefficients of all three forms.

For further computations, all persons taking recall are grouped together, indiscriminately, irrespective of method; and likewise for the recognition and recall. From these total distributions the total reliability coefficients and total average times of the table are computed.

Now, although for fifty questions, *i.e.*, *the same number of questions*, done by subjects on recall, recognition and true-false forms respectively in an average number of minutes of 6.9, 5.6, and 3.6, the recall method is the most reliable, it does not necessarily follow that the method is the most reliable for *the same average amount of time spent in examination.*

The average times show that, in the same time required to give one recall question, 1.23 recognition questions may be given and 1.92 true-false questions. The question then is, "Will the added nu nber of questions which may be given in the same time be sufficient to increase the reliability of either recognition or true-false to a point exceeding the reliability of the recall?" To answer this question theoretically, we make use of the usual formula for reliability coefficients of n forms of a test with n forms,

$$(1) \qquad r_{nn} = \frac{n\, r_{11}}{1 + (n-1) r_{11}}$$

in which, r_{nn} is the reliability to be calculated, r_{11} is the reliability of the halves of the test, and n is the number of multiples of the halves considered. Table II gives the r_{11} values, correlations of one 25-question form with another 25-question form.

When $n = 2$, we get the reliability of a 50-question form of the test with another 50-question set; *i.e.*, Form A with Form B of the same test. The formula then becomes

$$(2) \qquad r_{22}\ (\text{recall}) = \frac{2r_{11}}{1 + r_{11}}$$

Solving formula (2) directly for recall, and formulae (3) and (4) below, respectively for recognition and true-false; *i.e.*, taking n of such value that equal times will be spent in examination, such time as is required to give 50 recall questions, or 6.9 minutes, we obtain the last horizontal row of Table II.

$$(3) \qquad r_{22}\ (\text{recognition}) = \frac{2(1.23)\cdot r_{11}}{1 + [2(1.23) - 1]\cdot r_{11}}$$

Digitized by Google

TABLE II

COMPARISON OF RELIABILITY COEFFICIENTS OF THE RECALL, RECOGNITION AND TRUE-FALSE TESTS (ALL METHODS GROUPED TOGETHER)

	Recall	*Recognition*	*True-false*
Reliability (r_{11}) of halves. 124 cases. Two forms of 25 each.	.448	.385	.340
Reliability of two 50-question sets. Formula (2)	.618	.556	.507
Average time in minutes to do 50 questions	6.9	5.6	3.6
Number of questions per unit of recall time	1.00	1.23	1.92
Number of sets of 25 questions to get equal reliability of .618	2.000	2.602	3.140
Reliability of Form A with Form B when 6.9 minutes examination time is used	.618	.607	.664

(4) $$r_{22}\,(\text{true-false}) = \frac{2\,(1.92)\cdot r_{11}}{1 + [2(1.92) - 1]\cdot r_{11}}$$

This table yields the interesting result that, with all methods grouped together, with equal examination time, the true-false test is the most reliable, the recall less reliable, and recognition least reliable.

In order to obtain an idea of the dependence of reliability upon the amount of time taken in an examination, we may plot the three equations involving reliability coefficients and examination time. This has been done in Fig. 28, where the ordinates are r_{nn}, and abscissae n, or rather, n reduced to time in units of the recall time, which is used as a standard of comparison for the other two, using the scale, 1 $n' = \frac{1}{2}\,(6.9) = 3.45$ minutes, in which n' is the average time in minutes required to do 1 multiple of 25 recall information questions. The numbers of questions, Q, done in t minutes, are given respectively by the formulae:

(5) Recall, $Q = 7.25\,t.$

(6) Recognition, $Q = 8.83\,t.$

(7) True-false, $Q = 13.89\,t.$

Digitized by Google

These equations may be conveniently plotted as straight-line equations on the same graph, using coördinates on the right-hand margin of the graph. From the graph one can thus determine both the number of questions and the examination time required to yield any predetermined reliability coefficient. This graph holds good only for the range of ability represented by the test group of subjects and for the conditions of the experiment above outlined.

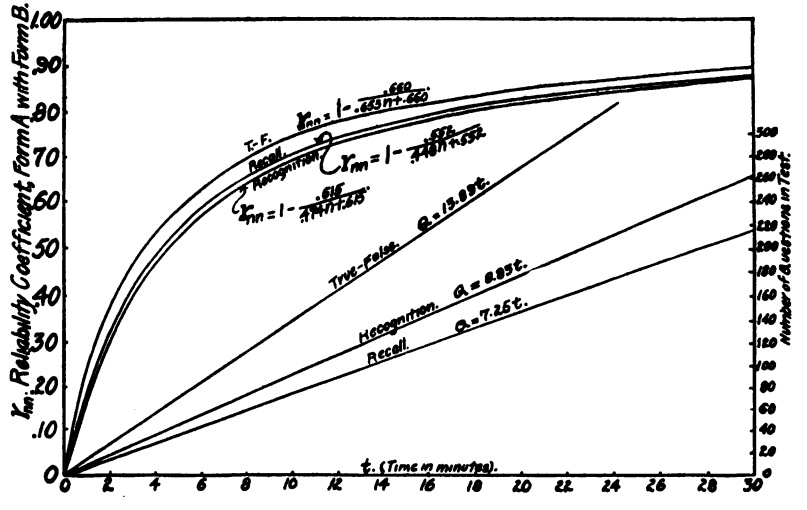


FIG. 28. THEORETICAL RELATIONSHIP OF RELIABILITY COEFFICIENTS TO EXAMINATION TIME.

There is some question whether grouping all persons regardless of method or order of giving the test may not obscure some of the differences which exist in the reliabilities. The average times, for instance, show that if a group of persons given the true-false or the recall or recognition first, followed by either of the other two, the group tested tends to take more time on the test taken first than any other group of persons given any other test first. Accordingly, the reliabilities have been computed, grouping together all the 76 people who took recall first, the 26 who took recognition first, and the 22 who took true-false first. Evidently the results on these tests are not influenced by previous consideration of any other form of test, and so correspond as closely as may be to an actual recall or other

Digitized by Google

test as given alone and without any other. The following table results:

	Recall	*Recognition*	*True-false*
r_{11} 25 questions	.743	.689	.561
r_{22} 50 questions Form A with Form B	.853	.816	.719
Average time in minutes	7.80	7.24	4.24
r_{aa} Equal examination time of 7.80 minutes	.853	.826	.824

The reliabilities here are much higher. The added "conscientiousness" when persons take the true-false first seems to make their average time, which is not increased proportionately, such that the true-false is only about on a par with the recognition form when equal amounts of examination time are taken. The conclusions to be drawn from this experiment are: (1) When equal numbers of information questions are given on the three forms of information test, the recall is always the most reliable, followed in order by recognition and true-false forms, but that (2) where equal amounts of examination time are taken on the three forms of test, the reliabilities do not differ greatly.

RELIABILITY OF VARIOUS TRADE TESTS

The above technique allows us to compare the reliability coefficients of various trade tests based on different numbers of questions. Reliability coefficients of the same magnitude are equal only when the range of ability is the same in the two cases being compared. Accordingly in Table III there is given a brief qualitative description of the group tested and their range of ability, as well as a brief qualitative description of the various tests, the number of questions in the test, the reliability of the halves of the test computed between the odd and even numbered question scores of the test, the reliability of Form A with Form B of the same test, $\frac{2\,r_{11}}{1 + r_{11}}$, and the reliability of Form A with Form B if fifty questions had been used in each test. The last column is thus the comparative measure

Digitized by Google

of the different tests, since it reduces all tests to the same number of questions. It is impossible with the data at hand to compute the reliabilities with equal amounts of examination time taken. The author wishes to emphasize the point that all correlation coefficients, of which reliability coefficients are but one variety, must be interpreted in terms of the range of ability; concretely, that since the range is limited to eighth grade students only, .795, the reliability of the girls' trade school extension students on the one-word-answer written switchboard test, is as high a reliability as would be represented by a much larger figure if several grades were to be included and the tests given to a larger range. Again, the best measure of an examination method is not the length of time spent in examination, but rather a composite of time spent in administration plus the time required for scoring. The time per person required for administration is often negligible as compared with the time required for scoring a test.

From Table III it will be noticed that the reliability of all trade tests, measures of acquired learning, is uniformly high save in the switchboard operators' test above mentioned, which is probably affected by the fact that only twenty questions were used in the examination, and in the test of progress in the university class in mental measurements likewise. The general information tests, ordinarily rated as tests of intelligence, are in a distinctly lower class as regards reliability.

This table presents the reliability coefficients of all the trade tests which the author has been able to collect up to the present time. Trade test methods give high reliability coefficients, provided enough questions, and sufficiently difficult or rather well scaled questions, are used. Forty selected difficult questions on the farmers' recognition test gave a higher reliability coefficient than one hundred and ten unrevised questions. Selected, or revised questions on the soldiers' mechanical test gave a higher reliability than the unrevised questions. The one-word-answer form of test generally gives higher reliabilities than the recognition or true-false forms.

SCHOOL EXAMINATIONS

Because of its high reliability and ease of administration and because it embodies all the superiority of the recall form of examination over the recognition form, the one-word-answer method is

Digitized by Google

TABLE III

Reliability Coefficients of Various Trade Tests

Group Tested	*No. of Cases*	*Range of Ability*	*Kind of Test*	*No. of Questions in test*	*r (odd-even)*	$\frac{2r_{11}}{1+r_{11}}$	r_{11} *basis of 50 questions, Form A with Form B*
Eyelet Machine Tenders	20	Three weeks to one year of experience	One - word - answer. Eyelet machine. Oral trade test.	50	.850	.919	.919
Bricklayers Standardisation Group	14	Apprentice to expert inclusive 4 A, 5 J, 5 E.	Army oral. One-word-answer. Army oral bricklayers test.	38	.929	.964	.985
Girls' Trade School Extension Students	14	All eighth grade pupils.	One - word - answer written. Switchboard oral test.	20	.436	.607	.795
Farmers and University Students	30	8 women novices. 12 men novices. 10 farmers.	Multiple choice (4 choices) written questions. Farmers' test.	"Easy" 70	.926	.962	.947
	30			"Hard" 40	.899	.947	.958
	30			Al 110	.957	.979	.953
Men and Women Summer School College Students	124	"Above college entrance requirements."	General information test. All methods combined.				
			Recall. One - word - answer.	50	.448	.618	.618
			Recognition of correct answer. 5 alternatives.	50	.385	.556	.556
			True-false.	50	.340	.507	.507
Men and Women College Students	61	Graduate students in college.	School test in mental measurements. True-false.	20	.25	.40	.625
	61		Same in one-word-answer form.	20	.50	.667	.833
Soldiers	271	Vocational courses E. & R. School.	Use of tools unrevised picture mechanical test.	46	.796	.886	.894
	223		Revised picture mechanical interest test.	45	.849	.920	.928

Digitized by Google

desirable in school examinations. The one-word-answer method can easily be adapted to school examinations, especially in those subjects which have a high information content: all types of science, physics, chemistry, botany, zoology, vocational courses, and technical subjects such as mathematics, statistics, mechanical drawing, logic. Any study which possesses a large number of terms peculiar to such studies, paralleled by "trade terms" in the trade, lends itself admirably to this method.

A physics examination, designed to illustrate the ease with which such questions may be written in a science subject, is herewith reproduced.

GENERAL PHYSICS EXAMINATION

To illustrate the trade test form of question

1. What do you call a minute particle of matter which is made up of atoms?
   Ans. Molecule.
2. What Physical property of iron allows it to be drawn into a wire?
   Ans. Ductility.
3. What do you call the product of the mass of a body multiplied by its acceleration?
   Ans. Force.
4. What physical property of a body tends to keep it in motion whence once set in motion?
   Ans. Inertia (momentum).
5. What do you call one-half the product of the mass of a body multiplied by the square of its velocity?
   Ans. Kinetic-energy.
6. What is the C-G-S unit of force?
   Ans. Dyne.
7. What is the diffusion of a fluid through a membrane called?
   Ans. Osmosis.
8. Who discovered the law of falling bodies?
   Ans. Newton.
9. How many feet per second of acceleration is given to a falling body by gravity?
   Ans. From 32.00 to 32.16.
10. What do you call a glass tube bent so as to drain water over a point higher than the surface of the water?
    Ans. Syphon.
11. How much is the weight of one cubic centimeter of water at its greatest density?
    Ans. 1-gram.
12. If the volume of a given weight of gas at constant temperature is doubled, how many per cent is the pressure decreased?
    Ans. 50.

Digitized by Google

13. What property of a tone is determined by the vibration rate?
Ans. Pitch.
14. What do you call it when rays of light are bent by going through a prism?
Ans. Refraction.
15. What shape of mirror reflects parallel rays of light to a focus?
Ans. Concave.
16. What are the first three colors of the primary spectrum, in order, at the long wave length end?
Ans. Red-orange-yellow.
17. At what Centigrade temperature does water freeze?
Ans. Zero (0).
18. What do you call the variation of a magnetic needle from a true north and south line?
Ans. Declination.
19. What kind of electricity is produced by friction?
Ans. Static.
20. What is the unit of electrical current called?
Ans. Ampere.
21. What other temperature scale is there besides the Fahrenheit and Centigrade scales?
Ans. Reaumur.
22. What do you call the force which makes a body fall to the earth when it is dropped?
Ans. Gravity.
23. What do you call a space from which most of the air has been pumped?
Ans. Vacuum.
24. What class of substances is generally easily volatilized?
Ans. Liquids (fluids).
25. What liquid is generally used in a thermometer for measuring temperature of more than forty degrees below zero?
Ans. Alcohol.
26. What do you call the ratio of the weight of a substance to the weight of the same volume of water?
Ans. Specific-gravity.
27. What is the unit of electrical power called?
Ans. Watt.
28. What is the unit of mechanical power in F.P.S. units called?
Ans. Foot-pound.
29. How many watt-hours in a kilo-watt-hour?
Ans. 1000.
30. What do you call the product of a force multiplied by the distance through which it moves?
Ans. Work.
31. What physical property has a body which causes it to rebound when dropped from a height upon a solid surface?
Ans. Elasticity.

Digitized by Google

32. What do you call it when a sound wave has been reflected from a building and comes back to you as a distinct word?
Ans. Echo.
33. What is the name of the hypothesis which says that no force is ever created anew or ever destroyed?
Ans. Conservation-of-energy.
34. What do you call the force which holds together two molecules?
Ans. Cohesion (adhesion).
35. What is the force called which holds a soap bubble together?
Ans. Surface tension (capillary attraction).
36. What do you call it when water takes up a quantity of gas?
Ans. Absorption.
37. What do you call that part of a body about which, if balanced, it would be free to revolve in any direction with equal ease?
Ans. Center-of-gravity.
38. What do you call half the swing of a pendulum's arc?
Ans. Amplitude.
39. How many foot-pounds in a horse-power?
Ans. 33000.
40. Who formulated the hydrostatic law which says that liquids transmit pressures equally in all directions and at right angles to the surface?
Ans. Pascal.
41. If the pressure on a block of wood two feet below the surface of water in a tank is one pound, how many pounds will it be if placed four feet below?
Ans. 2.
42. A block 2 cubic centimeters in volume weighs 8 grams in air; how many grams will it weigh in water at 4 degrees Centigrade?
Ans. 6.
43. How many feet will an ordinary suction pump lift the water?
Ans. From 27 to 32.
44. How many feet per second is the velocity of sound in air?
Ans. From 1080 to 1150.
45. If a string four feet long vibrates twenty times per second, how many times per second will a string eight feet long vibrate if under the same general other conditions?
Ans. 10.
46. What is the instrument called which is used with a prism to measure the wave length of light?
Ans. Spectroscope (spectrometer).
47. What do you call two antagonistic colors which produce gray when mixed on a color wheel?
Ans. Complementary (supplementary).
48. What do you call the force which makes the governor on a steam engine act?
Ans. Centrifugal.
49. What is the phenomenon called when one electric current in a wire sets up a current in a neighboring wire?
Ans. Induction.

Digitized by Google

50. What do you call it when hydrogen collects on the positive plate of a battery and thus stops the flow of current?
Ans. Polarization.
51. What is the decomposition of water by the electric current called?
Ans. Electrolysis.
52. What do you call the galvanometer which is used with an ordinary magnetic compass needle to measure electric current?
Ans. Tangent.
53. What kind of electricity is it called when electricity is generated by heating two different metals connected in series?
Ans. Thermo (thermo-electricity).
54. What do you call the ratio of the amount of heat required to raise a given weight of metal one degree Centigrade to that required to raise the same weight of water one degree?
Ans. Specific-heat.
55. What do you call the amount of heat required to raise the temperature of one pound of water one degree Fahrenheit?
Ans. British thermal unit (B. T. U.)
56. What, besides the volume of gas in a container, determines its pressure?
Ans. Temperature (heat) (warmth).

One may note that not only informational questions, but questions involving the application of principles can be easily adapted to this form of test. In physics the use of formulae as a means of expressing physical laws is very common. One aim of most physics courses is to teach the students to be able to solve these equations by the simple process of substitution in the equation of values determined from experiment, obtaining therefrom one unknown term of the equation. In an examination there is little merit in requiring the solution of equations which will involve decimal numbers. By using data which are in integral numbers and which are multiples of quantities which it may be necessary to divide, a test can be quickly made of the pupil's ability to apply the physical principles involved. Take, for instance, the problem, "What Fahrenheit temperature is 20° Centigrade equal to?" In order to transform 20° Centigrade into Fahrenheit temperature, the pupil has merely mentally to perform the multiplication of $\frac{9}{5} \times 20$ and then add 32°, giving an even quantity as the answer. It seems reasonable to suppose that the majority of students who can mentally perform that operation would also be able to perform the operation with the aid of a paper and pencil if 22° rather than 20° were the Centigrade temperature. Whether such is generally the case is not known. The relative value of the two

Digitized by Google

types of tests is one which can be solved only by the method of correlation. When evaluating the relative merits of the two kinds of problem one must also take into account the fact that possibly two or three questions of the trade test form can be given in the same time as one of the other form. The problem in the last analysis is then "Will n questions of the trade-test form, in the same amount of examination time, give us more or less knowledge of a pupil's ability in physics, than will $\frac{n}{3}$ questions requiring extended arithmetical operations to be performed with the necessary aid of paper and pencil?" Of course the use of paper and pencil in computing $9 \times 20^{\circ}$ should be allowed if the pupil needs them. It is also a question whether this mental type of computation should not be taught in physics. In many of our courses in science and mathematics it is possible that a great many problems which can be solved mentally or with little effort would result in clearer insight into the principles and the uses of the formulae involved, than would a few lengthy problems involving complicated arithmetical operations. Possibly this point is one of the main differences between ordinary school teaching and the training in the trade which a boy gets in industry itself. In industry a boy learns to do fractions by using a chart which converts the fractions into decimals. The commercial statistician, recognizing the fallibility of "long-hand" methods, uses a calculating machine.

Another point to be noticed in such examinations is the fact that in numerous instances the formula itself can be given.. This again is the practice followed in industry. Undoubtedly the most valuable commercial statistician is the one who can at a moment's notice write correctly any formula which he needs. The expert statistician, however, whenever complicated formulae are involved, resorts, as does the machinist in the machine shop, to his published handbooks of formulae, tables, and the like. A commercial physicist who could remember all his formulae would undoubtedly be able to do a larger amount of work in a day than one who had continually to resort to books, especially if the formulae were to be found in numerous books rather than in one hand-book. Such attempts to justify imposing routine upon a pupil in the study of any science are unworthy attempts at an outworn theory of discipline. They belong to the old puritanic philosophy of "discipline for discipline's sake, work for

Digitized by Google

work's sake," and do not fit in with the modern philosophy of letting a machine do the work and reserving the more costly human machine for work which requires planning and executive ability.

One of the great barriers to securing large amounts of data on school practice, as a basis for revising school administration and school procedure, is the large amount of time required to give, score, and evaluate the tests. The trade test form of question, if it can be effectively applied to school subjects, has much to be said in its favor on this score. The above physics test clearly indicates the possibility of quickly examining a student on a very great many elements of the course pursued. Testing on many elements is a much more reliable method of securing an accurate evaluation of the individual's ability than testing on only a very limited number, as is done where a composition is written in response to each question of the examination. The value of the one-word-answer method in educational work remains of course to be proved. On the point of time taken to give the test and to score the results, we may state at the outset that a very great advantage can be secured, since the tests can be quickly scored by stencil.

The use of the underscoring, or multiple choice method, of giving an examination may be considered on the basis of ease of scoring. This method is even easier to score than the one-word-answer method. The value of a test, however, should not be sacrificed merely for the ease of scoring. Using the underscore method, a pupil cannot make use of the independent thought which he must use when answering the recall or trade test form of question. Many more items of information can be touched upon by using the underscore method than can be tested in the same time with the one-word-answer form of question. Certain types of information may be easier to present with the one type of examination than with the other.

Another point of practical consideration is that the one-word-answer form of question cannot be used very often in the recitation itself. It is difficult to form good trade test questions merely on the spur of the moment, and the same difficulty as is encountered in the trades applies here also. If only seventy-five good trade questions can ordinarily be discovered in a trade, then certainly not that number will be found in the ordinary amount of subject matter covered by a class in physics in a single recitation. It is possible to multiply

Digitized by Google

examples of trade test applications of formulae, for instance, and so occasionally to obviate this difficulty.

Care should be take in using the trade test form of question that the teaching be not made to conform to the arbitrary standard set up by the trade test. Teaching would become rather artificial if the teacher always stressed the child's giving the short answer form of replies which the trade test demands. Teaching would become lifeless if the aim of the teacher were directed towards merely preparing the pupils to answer trade test questions. This applies, however, with equal force to any measure of an educational product; the teaching always will become very lifeless if the teaching is directed towards preparing pupils merely to pass tests. There are some obvious exceptions, however; in the case of arithmetic or any such mathematics as the formulae of physics, the educational aim should certainly be to enable the pupil to use such fundamental operations, to solve equations with ease; in other words, to use his acquired skill in exactly the same manner as is required in the trade test question.

The one-word-answer form of question can be used to stimulate the pupil to definiteness of expression. Both the composition type of examination question and the trade test form should be used, perhaps alternating the one with the other. The former encourages clarity of expression in composition; the latter, definiteness. The latter type of ability is perhaps demanded in industry to a greater extent than the former. The leading question, a type of question which indefinitely suggests the line of approach of the student, will continue to be used by the teacher as a means to stimulate the child's curiosity and lead him on to rely on his own ability. The picture trade test method, essentially an oral method, is probably suggestive of some possible school examination methods such as have hitherto not been attempted. The picture lathe hand test shown on pages 8–14,if made easier than it is now, could be adapted to written examinations in that subject in trade schools. The picture form of examination will also readily adapt itself to the teaching of school subjects.

Among the specific criticisms of the use of the one-word-answer, the picture, or the multiple choice tests in examining in school trade subjects is the often recurring one that such tests are tests of information mainly and do not require independent thought and organization of subject matter. It is true that, in any trade examination

Digitized by Google

made up on these methods, a casual inspection will reveal a seeming preponderance of "informational" questions. This objection is minimized by the consideration that the importance of "thinking" in doing many of the routine tasks of industry has probably been overestimated. The most successful army oral questions were found to be the questions which seemingly best sampled the everyday habits and knowledge of the tradesman. In present-day industry, the workman's task and the directions for performing it are worked out for him in advance; he relies largely on charts and mechanical devices for such "figuring" as is required; moreover, methods of scientific management and specialization in industry are continually breaking up trades into series of jobs, each of which is so simple as to be performed adequately by persons of only fair intellectual or mechanical ability.

Finally, the whole question of whether or not many informational questions are desirable is one which will ultimately be settled only by the correlation resulting from experiments in which both "informational" and "thought provoking tests" are tried with the same people. It must be remembered that thinking takes much time, while writing immediately recalled information takes much less time *per question*. The question then is, "For a given amount of examination time, will X thought provoking questions or $(X + Y)$ informational questions give higher correlations with demonstrated ability in performance on the job?" These methods are very apparently of limited value in the recitation, upon which should be thrown the chief burden of teaching pupils to think clearly.

Psychologically, there is no reason why the results of a complicated thought process cannot be indicated by a simple one-word-answer response on the part of the subject. The mental arithmetic "reasoning" problems, to be found in great abundance in any elementary arithmetic text, generally fulfill the major requirements of a good one-word-answer question. Few people would maintain that these are not thought provoking. It merely requires more skill on the part of the question maker to write "thought provoking" questions than to write "informational" questions.

All these methods are subject to the danger of preparation by coaching. The one-word-answer form of test is probably more subject to this difficulty than the picture form. Certain forms of psychological tests are equally subject to this danger. The obvious

Digitized by Google

solution to the difficulty lies in having so many questions of each of the different forms that alternative tests may be provided.

The tradesman in industry has seldom had an incentive for coaching up on any form of trade examination; in trade school procedure, examinations are expected as a matter of course. In the trade school the same material may readily be formulated in time into four forms of test, the recall one-word-answer, recognition or multiple choice, true-false, and picture, in order to overcome undue influence of "coaching for examination." The coachability of a test is probably in inverse proportion to the number of thought provoking questions included, obeying thus the same law as applies in general to all examination methods.

Finally we may dismiss with but little discussion the possible objection that the acquisition of trade knowledge is not the aim of the vocational school. The present avowed aim is the preparation of youth to earn a livelihood in the vocations. We have before indicated the growing need in industry of the workman possessed of a goodly number of specific habits, skills and knowledges, and the growing tendency to simplify trades to such an extent that less intelligence and less "mechanical ability" is required for success on the job by the majority of workers.

Digitized by Google

CHAPTER III

TESTS OF TRADE CAPACITY OR PROMISE

The possibility of predicting a person's probable success at a job depends upon a relatively fixed condition of the job, a knowledge of the requirements of the job in the way of qualifications of its workers, and methods of measuring human qualifications and evaluating these in terms of degrees of success. Historically there have been three stages in the methods of attacking the problem, each of which has been characterized by developments in the methods of measuring the job on the one hand, and the worker on the other.

In the first stage, which we may call the "mental type" stage, it was believed that success at a given job was possible only for persons of a given mental type. Efforts were accordingly directed towards determining the mental types best suited to given jobs, and toward determining to what mental type a given person belonged.

This stage gave way to a second stage in which was attempted an analysis of the lives of successful and unsuccessful men, in order to discover what psychological or physiological trait or traits might be responsible for their success or lack of success. These methods were largely a failure. Statistically, the difficulty with them is that any one such single trait is very likely to have only a very low correlation with job success. On the side of measuring the job, efforts were made to classify jobs according to the predominant traits of character and mental ability necessary for success on the job.

This second stage then merged into the third, in which, not single traits, but combinations of traits, are considered responsible for the degree of success on the job. The development of tests of general intelligence marks the turning point in this stage. The general intelligence tests were found to be good predicters of general school success of school children, and to be of value in detecting children of abnormal mentality—sub-normal or super-normal ability to profit by the instruction received. From this it was but a step to infer that general intelligence scores would be found very significant in placing persons properly in industry. We are at present in this third stage but rapidly entering a fourth, in which the emphasis will be placed upon the development of statistical methods for the

Digitized by Google

evaluating into one composite score of fitness for the job all varieties of facts about an individual: his intelligence, parentage, physical capacity, formal education, informal trade training in mechanical manipulation of his environment and so on. Evidence has been accumulating rapidly in the last few years to point to the fact that individual traits may singly correlate but slightly with job success, but that a composite of such traits may correlate rather highly with job success. The method of partial correlation enables one to make the best possible mathematical prediction (assuming linearity of regression lines) of job success from a number of variable measures or tests. The statistical requirements for a composite test of high predictive value are that the correlations of the tests with the criterion of success be high in relation to the intercorrelations among the tests themselves. Furthermore, there is coming to be stressed the fact that the reported correlation of the composite test score with the independent measure of job success can only be interpreted in the light of the range of ability covered by the test subjects; or, concretely, that a correlation of .60, where only eighth grade children are involved, indicates as good a composite test as would be shown by a much higher correlation if children from all the grades were included.

On the side of measuring the job, there has arisen, just at the transition point from the third to the fourth stage of the development of human measurement, the method of job analysis. In its present best developed stage, the job is described in terms of concrete job processes or equally concrete and measurable human qualifications. Already we have heard hints of the possibilities of development of methods of measuring jobs by a scale of composite comparable measures for all jobs, similar to the scales already being developed for human measurement. Fitting the worker to the job is only half of the larger social problem involved in vocational guidance. In many cases, jobs should be modified to suit workers. Such modification will tend to come about automatically as soon as the job analysis has caught up, in its process of development, with human measurement and entered likewise into the fourth stage.

The names of many experimenters are connected with the development of present methods of predicting school success by the use of composite scales. The application of like methods in predicting industrial success has followed the former, and hence has fewer

Digitized by Google

names associated with it. Mention might be made of the names of Thorndike, Scott, Thurstone, Link, Burt, Marcus, and Otis.

Some six years ago Dr. Thorndike devised and put into use a series of intelligence tests for the selection of clerical workers for the Metropolitan Life Insurance Company. Among the tests were tests not only of general intelligence but of abilities similar to those employed by clerical workers in their work. The examination, constructed in alternative forms in order as far as possible to overcome coaching difficulties, is reported to be giving successful results after six years of trial, and yielding correlations with job success of probably over .65.

Dr. Scott has devised for a number of firms composite scales for the selection of salesmen and clerical workers. His tests have often included the use of some form of group judgment or rating scale of important character qualities presumably not tested by the psychological tests.

Dr. Thurstone has developed a widely used test for predicting clerical ability. During the war, his researches into the qualifications required for telegraphers involved physiological or psycho-physical tests as well as psychological tests.

Dr. Link [1] reports many correlations sufficiently high to justify the belief that composite scales of various tests, selected on a "trial and error" basis, may be used to predict the chance of success in many industrial jobs, chiefly of a repetitive motor or manual type.

Dr. Burt [2] reports on the success of weighting tests by the method of partial correlation, using different weightings of the same tests for different jobs, by means of which the probable degree of success of a given worker in the rubber industry may be predicted for a number of jobs. Essentially the same technique was developed by Dr. Thorndike during the war, for rating members of the S. A. T. C. for different branches of the army.

Mr. Marcus [3] has shown that a composite of psychological tests may be more efficient in predicting the probable success of certain

[1] Link, H. C., *Employment Psychology*, 1919.

[2] Burt, H. E., "Employment Psychology in the Rubber Industry," *Journal of Applied Psychology*, Vol. 4, No. 1, pp. 1–7.

[3] Marcus, L., "Vocational Selection for Specialized Tasks," *Journal of Applied Psychology*, Vol 4, Nos. 2 and 3, pp. 186-201.

Digitized by Google

manual clerical workers than the ordinary civil service examinations given these workers.

Dr. Otis [4] has devised a clerical test which he reports as being valuable in predicting success of clerical workers in a large silk mill. He also reports "zero" correlations between a composite score of "intelligence" tests and performance on the job of manual workers in the same silk mill. He thus concludes that intelligence is not an important consideration in success in such jobs. A number of attenuating factors enter into his results.

With all the above developments in the methods of measuring human fitness for jobs, more time should be devoted in the future to methods of measuring jobs. The lines of development in job analysis will be towards:

1. Securing greater definitness of descriptions, as measures of the job which can be verified by other competent persons.
2. Determining methods of comparing the relative merits of one job with another, with the practical point of view of eliminating the discrepancies in wages and working conditions of jobs which require approximately the same composite of talents in their workers.
3. Rating jobs on scales of physiological, psychological, and sociological qualifications of workers.
4. Providing for a series of gradations of preferences for jobs by workers on the grounds of monotony, cleanliness, and all those job conditions which affect the emotional life of the worker.
5. Providing for formal methods of training talent for the job among the workers, and of rewarding periodically and automatically any acquisition of better qualifications for the job by job promotion, either in wages, responsibility or social recognition.

JOB ANALYSIS

The term "job analysis" is used rather loosely. By it is meant, as we have more or less implied above, the entire system of industrial and personnel research, sometimes referred to as a "labor audit." In its narrowest sense, job analysis, made primarily for the purpose of hiring and training, transferring or promoting employees, is the analytical process of acquiring industrial facts. This is to be carefully distinguished from the products of the analytical process, such

[4] Otis, A. S., "The Selection of Mill Workers with Mental Tests," *Journal of Applied Psychology*, Vol 4, No. 4, pp. 339-341.

Digitized by Google

as job specifications and personnel specifications, which are written reports of the work done by such analytical methods.[1]

Job specifications are formulated groups of facts which relate to the nature and conditions of the work, the duties of the worker and certain conditions of service, such as pay, hours and promotions. The personnel specifications present the same type of facts for the worker, for example, his qualifications, allowable handicaps, and required training. Only as an abstract logical analysis are the personnel specifications to be thought of independently of the job specifications. Both must be developed at the same time; both are logically used together in an attempt to fit the job and the worker to each other.

Every job description should contain at least a minimum amount of description of the various operations, tools, problems and the like, expected of the worker in the job. In addition, a condensed personnel classification should be included if there are any special requirements on the part of the worker which *he himself* can easily and objectively measure to see if he fits up with the demands of the job. Such, for instance, is the amount of education required.

As a part of the job specifications, whether or not included in the final typewritten or printed job specifications, or whether given on an independent new worker initiation sheet, there should be included all the information which the new applicant will want to have about the job outside of the duties of the job itself: the minimum and maximum wage to be expected, the piece work rates, the average length of time required by workers of average intelligence to learn the job, the average length of time spent at the job before promotion, the working hours, statements in regard to holidays and vacations, and social, educational, industrial or other privileges which may be granted the worker.

There should be collected other items which have a bearing upon the applicant's contentedness while on the job, such, for instance, as whether or not the worker is required to sit or stand most of the time on the job, whether the work is dusty, hot, damp or otherwise disagreeable, by what type of men the job is manned, referring to such things as nationality and color.

[1] An excellent summary of all job analysis work to date is given in Meine, F., *Job Specifications*, Federal Board for Vocational Education, Employment Management Series, No. 3, Bulletin 45, 1919. See also Tead, Ordway, *The Labor Audit*, Employment Management Series No. 8, Bulletin 43, 1920.

Digitized by Google

In the personnel specifications should be included the degree of intelligence and the amount of education, experience or technical knowledge required on the part of the applicant, and factors regarding the work itself, translated into human qualities, such as the height, the amount of strength, the quality of sight and hearing, and other *measurable* physical characteristics that may be required on the part of the applicant.

The utmost care should be taken to see that whatever goes into the job and personnel specifications, should consist of items, which can be measured with some reasonable degree of accuracy by the people who will have to use them. Too frequently have there been long, wordy descriptions of the job as being one which required "speed, endurance, industry, skillful manipulation, etc., etc." The analysis can thus be worded in such skillful manner that even a sophisticated job analysis expert will be misled by the fluency of the phrases. The principle always to be followed is simple: Whenever tempted to use a term like "skillful manipulation," the job analyst should always ask himself, "*How much* skillful manipulation is necessary"; and "Can the person who is to use the job specifications in any way measure this same trait."

The following questionnaire, only a beginning in this field rather than a completed product, will suggest some of the minimum items of information which it may be worth while to collect in the machine-like trades. This questionnaire is designed for a preliminary general job analysis with the expectation of later following it with specialized surveys of sanitation, training and psychological investigation. Note that as many of the items as possible are in the form of specific answers to specific questions for underlining and that space is always left for necessary unforseen classifications. Such items as have no claim to objectivity may be generally recognized by having lines in place of descriptive answers to be underscored. Notice also the emphasis upon learning just where and how all available records of the job and workers are kept.

GENERAL JOB ANALYSIS QUESTIONNAIRE SHEET

Name of Trade..... (General) (Specific) Symbol..... Present Code Number.....

Reported by..... Date..... Dept..... Foreman.....

Digitized by Google

I. THE JOB.

A. Analysis of the Job in the scheme of:

1. Classification of jobs:.....
2. Products produced by worker:.....
3. Machine (Pictures of all types of machine) Types are:.....
.....
4. Number of workers required per production unit:.....Unit is:.....
5. Number of workers hired in this occupation in this department on this date:.....

B. Similar Occupations:.....

C. Substitute Occupations:.....

D. Physical Conditions of the Job:

1. Place where work is done:	MACHINE	DESK	FLOOR
	BENCH		
2. Posture:	STANDING	SITTING	STOOPING
	CLIMBING	WALKING	LIFTING
	PULLING	SHOVING.....	
3. Activity:	WALKING	LIFTING	PULLING
	SHOVING	CLIMBING	KICKING
	FINGER-MOVEMENT		
	WRIST-MOVEMENT		
4. Physical Conditions:	WET	DRY	DUSTY
	OILY Fumes of.....		VAPOR.
	LIVE-STEAM	HOT	COLD
	NOISY	DIRTY	STICKY
	CLAMMY	ELECTRICAL.....	
5. Lighting:	Source.....Location.....		
	Intensity.....Glare.....Shadows.....		
6. Cutting or Lubricating fluids:	What?.....When used?.....		
	.....		
7. Feeds:	MACHINE-FEED	HAND-FEED	SEMI-AUTOMATIC.....

a. Is speed independent of the operator? YES NO

b. What is mechanical maximum? (Explain).....
.....

Digitized by Google

8. Rythm: REPETITIVE NON-REPETITIVE
   a. Is raw material uniform? YES NO (Explain).....
   b. Is each operation uniform? YES NO (Explain).....

9. Monotony:.....

10. Motion Study (chronological sequence of operations. Describe briefly).....
.....
.....
.....

11. Fatigue:
   a. Degree.....Type.....
   b. No. of cycle repetitions per HOUR.....
   c. Size of unit of production.....
   d. Amount of lifting done.....Weight of piece handled.....
   e. Estimated number of footpounds of work done per day.....

12. Members used:
   a. BOTH HANDS LEFT HAND RIGHT HAND
   b. BOTH LEGS LEFT LEG RIGHT LEG

E. Miscellaneous Conditions of Job:
1. Does worker have to supply any tools, clothing, etc? YES NO
   Specify:.....
2. Chairs, foot rests, back rests, arm rests, coat racks, eating places:
.....
3. Impediments of hair, clothing, rings, jewelry, etc.:.....
.....
4. Vacations:.....
5. Holidays: FOUR REGULAR.....
6. Night work: Amount.....Rate.....Hours worked.....
7. Overtime: Amount.....Rate.....
8. Sunday work:.....
9. Rest periods:.....
10. Lunch interval:.....
   a. What do workers do during lunch interval?.....
11. Deductions, deposits, fines:.....
12. Tardiness: Amt.....Records of, (where and how kept?).....
13. Absences: Amt.....Records of, (where and how kept?).....

Digitized by Google

14. Are workers allowed to talk to each other during work? (describe):
.....

15. Do mechanical factors of job preclude possibility of same production with shorter hours? Foreman's opinion..... Interviewer's opinion
.....

F. Accident and Health Hazards:

1. Accident risk (nature and causes):.....
.....
2. Percentage of accidents to workers: (Foreman's opinion).....
3. Safety devices and use:.....
4. Attention to job required to prevent accident:.....
5. Liability to accident through carelessness of others:.....
6. Health hazards (nature and causes):.....
.....
7. Occupational diseases:.....

G. Inspection:

1. How often is product inspected?.....By whom?.....
2. Type of inspection: LIMIT-GAGE MICROMETER.....
3. Responsibility for inspection:.....
4. Are finished products turned back on account of inspection? (Describe).....
5. How often are machine and tools inspected?.....
By whom?.....
6. Spoiled work: Amount.....Cost.....
Cause: HUMAN MACHINE.....
7. Are records of spoiled work kept? YES NO (Describe).....
.....
8. Can spoiled work be traced to the operator producing it? YES NO (Describe).....

H. Production, Turnover, Security of Job:

1. Regularity of work: seasonal.....lay-offs.....
2. What are grounds for discharge of worker?.....
3. Production:
   a. Average production of workers on job: (Foreman's estimate)
   .....
   b. Minimum:.....Maximum:.....
4. Foreman's opinion as to whether greater production possible and desirable:.....
.....

Digitized by Google

5. Production asked by Planning Dept.:.....
6. Interchangeability of tools and machines, idle machines:.....
.....
7. Rush orders (nature):.....
8. *a.* Is worker idle while setting up machinery, tool setting, inspection, repairs? (Describe).....
b. Does he rest in such intervals? (Describe).....
9. Labor turnover on this job (Foreman's estimate):.....
a. Average length of employment of worker on this job. (Foreman's estimate).....
b. If men and women are both employed, which stay longer? (Foreman's estimate).....
10. *a.* Is production adaptable to count YES NO (Describe).....
.....
b. Are production records kept? YES NO Where?.....
How accurate?.....
11. Average length of training for average worker to become proficient on job.....
12. Cost of labor turnover:.....

I. Labor Supply:

1. Any apprentices (understudies) being trained?.....
2. Deficiencies in qualifications of workers (Foreman):.....
.....
.....
3. Will supply meet demand in future? (Foreman) YES NO (Describe).....
4. Source of supply:.....
5. Transfers: Amount.....Per cent formed of turnover.....
6. How are workers selected (function of foreman).....
.....
.....

II. THE WORKER:

J. Description of Duties:

General duties (Describe briefly):.....
.....
.....
.....
.....

Digitized by Google

Special duties (if sub-group of larger group, e.g., dial press tool setter vs. blanking gang press tool setter):.....

.....

Miscellaneous:

1. Set tools:.....
2. Oiling:.....
3. Belts:.....
4. Brakes:.....
5. Sharpen tools:.....
6. Repairing:.....
7. Inspecting:.....
8. Teaching:.....
9. Safety:.....
10. Responsibility:.....

K. Qualifications:

Previous trade training required:

1. Blue prints:.....Drawings:.....
2. Order tools:.....
3. Measuring tools:.....
4. Names of tools, machines, parts:.....
5. Knowledge of materials (qualities and quantities):.....
6. Teaching:.....
7. Inspection:.....

Requirements of schooling, intelligence, training on the job.

1. Is the job: SKILLED UNSKILLED.....
2. SCHOOLING:.....
3. Technical training:.....
4. Deficiencies in TRAINING of workers (Interviewer):.....

   .....
5. Necessary previous experience in same or allied job:.....

   .....What job?.....
6. Average length of time for average green hand to come up to average piece rate earnings:.....
7. Are motions, methods, feeds standardized?.....
8. Qualifications for a worker to be trained (reasons):.....

   .....
9. Intelligence:.....
10. Training on job: amount.....By whom given?.....

Digitized by Google

11. How introduced into new job?.....
12. Is new worker followed up in ten days? YES NO.....
13. Any written or oral instructions to new employee? YES NO
    By whom?.....What?.....
14. Required ability in mathematics: COPY FIGURES, ADD, SUBTRACT, MULTIPLY, DIVIDE, DECIMALS, SQUARE ROOT, SOLVE FORMULAE, CALCULUS.
15. Nationality preferred (foreman's reasons):.....
16. Literacy: .....GIVE-ORDERS READ ORDERS UNDERSTAND-VERBAL-ORDERS.....

Physical:

1. Eyesight: acuity.....glasses an objection.....color .....
2. Hearing: acuity.....
3. Sex: M F (Reasons).....
4. Height:.....
5. Weight:.....
6. Arms: length.....strength.....
7. Fingers: length.....dexterity.....size.....
8. Fingernails: toughness.....
   Wear fingerstalls? YES NO
9. General strength:.....
10. Strength of back:.....Lifting.....
11. Endurance:.....
12. Agility in moving about:.....
13. Sensitiveness: touch.....pressure.....weight.....
14. Attention required:.....
15. Personal appearance:.....
16. Physical defects disqualifying:.....
17. Physical defects allowable:.....
18. Can women do the job? YES NO.....Ever tried? YES NO
    Foreman's opinion of results:.....
19. Can cripple, aged, or physical defective do the job? YES NO
    Remarks:.....
20. Can mental defective do the job? YES NO Remarks:.....
21. Can left-handed person do the job? YES NO Remarks:.....

L. Wages and Promotion:

1. Line of promotion:.....
2. Do workers know line of promotion?.....
3. Do promotions come unasked for?.....

Digitized by Google

4. Any definite plan of promotion?.....
5. Wages:

Basis of Pay	*Lowest*	*Highest*	*Average:*
Day rate	.....	.....	.....
Piece rate	.....	.....	.....
Contingent rate	.....	.....	.....

6. Nature of contingent rate:
   a. On day rate:.....
   b. On piece rate standard:.....
   c. On piece rate earnings:.....
7. Average daily earnings:.....
8. Average weekly earnings:.....
9. Average yearly earnings:.....
10. Number of all on this job receiving piece rates:.....
11. How long before green hand put on piece rates?.....
12. Stereotyped wages in production:.....
13. Can operator trained on one type of machine operate another.....
14. Does this job prepare operator for higher job?.....
15. Are wages dependent upon: WORKER'S-OWN-ABILITY GROUP-SPIRIT.....

.....

.....

.....

.....

It will be noted that much information is of the purely subjective type, as, for instance, "attention required, degree of fatigue, etc., etc." At the present time but little better can be done than these subjective judgments which approach dangerously close to the type of character analysis which we have criticized so severely. It is desirable to find objective methods for expressing these items of a job. It would seem possible to devise a very simple objective rating scale for monotony, for instance:

DEGREE	DEFINITION
1	*as monotonous as a dial press job.*
2	*as monotonous as a production lathe job.*
3	*as monotonous as a criminal lawyer's job.*

Digitized by Google

Such a scale with intermediate degrees, carefully and thoughtfully worked out, would yield results by means of which various jobs might be semi-objectively compared for monotony. Outside of being valuable in turnover investigations, such information is of little or no value in hiring, since no one is competent and no test method is available which will predict with surety in advance whether an applicant will work contentedly or not at monotonous work. If, however, an applicant should report to the interviewing examiner that he had worked steadily for X years at a job Y, which the examiner knows is a job with degree of monotony 1½, then he would have some assurance that the applicant would be satisfied on a similar job Z with the same degree of monotony. Men do not fall into two clearly defined groups, the honest or the dishonest, but are of many degrees of honesty; jobs likewise are not monotonous or varied, but of all degrees of monotony. Furthermore, the job analyst must not fail to make the subtle distinction that a job, monotonous to himself to the extent T, may have a lesser monotony to the worker who works at that job as represented by the expression, $T - K$.

There is unlimited room for improvement in the methods of measuring jobs. This statement does not imply that there is not also room for improvement of tests used in measuring human capacities. One line of development, as yet but little used, is that of measuring by standardized tests a worker's performance on the job. In the training school of industry this aspect of human measurement would take the line of measurement of product turned out after a given amount of trade instruction on standard test jobs of production; in the trade school, it will mean performance tests, of a standardized nature, given to the students in trade courses to measure their ability in the course taken; while in the vocational guidance bureau, it will mean subjecting boys and girls to trade instruction with the sole purpose of testing their ability to learn the trade processes taught. It has been the privilege of the author to become acquainted with an educational experiment which combines both the latter purposes in a "trade extension school for vocational guidance."

EVALUATION OF VOCATIONAL GUIDANCE TESTS

At the Manhattan Trade School for Girls, New York City, there is being conducted one of the most interesting experiments to be found anywhere in this country in extension education and vocational

Digitized by Google

guidance by means of tests. The uniqueness of the plan, the numbers of pupils given advice, and the fact that tests are being used actually in basing action upon them, rather than as mere interesting but unused researches, entitles this extension department to an enviable position among the foremost vocational guidance programs to be found to-day. Every second week throughout the school year, a new group of from 100 to 120 eighth grade public school girls from various public schools of New York City come to this extension department for a two-weeks course. The aim of the department is both to instruct the pupils on various vocational opportunities and to test by means of performance tests their ability to learn a few fundamental things about a limited number of these trades or vocations.

The teachers of this extension department make occasional visits to the industries, working on follow-up reports of girls placed in industry by the school, and thus become acquainted with the facts of the various industries touched upon in the extension school course. Thus information becomes available in regard to average wages, range of wages, trade opportunities, active and dull seasons, sanitary and other working conditions of the various trades. These facts are imparted to the girls by the respective commercial and industrial teachers. The information given on commercial work covers information on the many subdivisions of commercial work; for instance, the positions of typist, stenographer, secretary, telephone operator, computing machine operator, filing clerk, bookkeeper, etc.

In addition to this work, educative and highly desirable as a part of the vocational guidance program, the larger part of the two weeks time is given over to the performance tests. The general plan of the tests is first to give oral instructions in the method of doing a fundamental operation, and then to judge *subjectively* the excellence of the *objective* product produced.

The pupils are rated by the teachers at approximately the end of every hour's work during the fifty hours spent in this extension school. In some cases it is possible to secure independent ratings of two different teachers of the same industrial subject, hand sewing, for instance. This makes for more reliable ratings when the double ratings are possible. The great amount of rating required necessitates that the teacher become acquainted as rapidly as possible with the pupils. What might to the outside observer appear as an excessive amount of rating is therefore very valuable in insuring that

Digitized by Google

judgments will be made in as few cases as possible without knowing pupils. The large amount of rating required means that the time that could be devoted to objective scoring of products is very limited; in fact, to the extent that at the present time no objective scoring is done. Objective scoring, using scales, tests scored by stencil or otherwise, could be adopted only by devoting to such work time now spent in subjective scoring.

As a final record of each pupil, a qualification card is made out in duplicate. One copy of this card is retained in the files of the extension school for possible later reference or follow-up work; the other is sent to the principal of the school from which the pupils came, to to be used by him in supplementing his own guidance given to graduating pupils. An evaluation of the items of this qualification card is here undertaken in an effort to determine the value of the several tests, both theoretically and practically, with the hope of suggesting improvements in the present tests or test technique, with the practical point in view of securing (*a*) better predictions of after-school sucess in the vocations undertaken, and (*b*) greater ease and objectiveness of scoring the various tests in reporting the results.

The scores on the qualification card are not the scores received on the several tests; the hand sewing scores of the qualification cards, for instance, represent the results of the first subjective summation of scores in the four abilities of hand sewing, "accuracy, speed, neatness, handling," in making a simple hem handkerchief and a child's cap; the pasting scores of the qualification card represent the first subjective summation of a series of scores of the four abilities of pasting "accuracy, speed, neatness, handling" in doing sample mounting of colored cloth triangles upon a cardboard backing and making a paper covered pasteboard box with cloth reinforced corners.

The test scores of qualification cards are subdivided into commercial tests, ten in number, and industrial tests, nine in number. To evaluate these, a celluloid stencil was constructed, on which each test and each score or rating on a given test was arbitrarily given a numerical value. It is only by assigning numerical values to "high average" position of check mark on a line opposite the English test, that such scores can be effectively dealt with statistically.

The commercial tests scores were all reported as "good, fair," and "poor," finer differentiations of ability being secured by significance

Digitized by Google

being attached to the position of the check mark, indicating the score in the test, upon a ruled line approximately 2.6 inches in length. On the celluloid scoring stencil, this length was subdivided into nine equal parts, numbered from 1 to 9, in increasing order of merit of score obtained. The position of the check mark as falling in one of these nine vertical columns of the stencil is the score in each of the commercial tests, the variables being numbered 5–14, inclusive. These check mark estimates were subjectively evaluated and a commercial "stamp," or recommendation of the director, placed on the card as variable No. 15. The various stamps and their evaluation, as well as the similar industrial "stamps," are explained below. The industrial tests, in their various subdivisions, had their scores reported as "good, fair" and "poor," arbitrarily given numerical values respectively of 3, 2, and 1, in variables numbered 16 to 34, inclusive.

The several tests, briefly described, are as follows:

I. Commercial Tests.

Variable 5. English. Three compositions and a final test are required in the course:

A. Composition: Skilled and Unskilled Labor.
B. Composition: Vocations for Women.
C. Composition: Effect of the Introduction of Machinery.
D. Final test.
   a. Dictation: Discourse between two people.
   b. Composition: What These Courses in the Extension Classes have Meant to Me.

Time:

Lectures, preliminary to compositions, 30 minutes each.
Compositions, 30 minutes each.
Final test, 60 minutes.

Variable 6. Arithmetic.
Single problems in arithmetical operations.

Variable 7. Penmanship.
Based on penmanship on written work of compositions.

Variable 8. Switchboard.
A series of demonstration performance lessons on the use of a private branch exchange switchboard, the various pupils taking turns in being, "private branch operator," "lady in a residence," "desk clerk," etc., being rated by the teacher in each capacity in turn.

Variable 9. Filing.
A series of three lessons with tests on each.

Digitized by Google

A. Alphabetical filing. Test of arranging 50 cards in alphabetical order and copying the order on paper.
B. Geographical filing. Test on same.
C. Telephone Directory and its use. Test of using directory in locating designated names.

Variable 10. Stenography.
A series of lessons, beginning with the single strokes and leading up to the final construction of simple words and sentences. Ability scored on the basis of reactions to simple dictation involving beginning principles of stenography.

Variable 11. Typewriting.
A series of five lessons leading up to the final formation of simple letter groups, or syllables, by the touch system. Practice sheets are preserved and subjectively scored by the teacher.

Variable 12. General Adaptability.
A subjective judgment of presumable general intelligence.

Variable 13. Personality.
A subjective judgment.

Variable 14. Appearance.
A subjective judgment. Stressing neatness rather than personal beauty.

Variable 15. Commercial Estimate, or "stamp."
This is the director's subjective summation of all the check marks of the above tests. This is the recommendation of action to the principal of the school from which the pupil comes. A series of rubber stamps are used, an imprint of one of which is affixed to each card. As arbitrarily evaluated, after comparison of meanings of the various ratings of the various teachers by drawing on a ten-inch line the division points of the various meanings, the stamps were given numerical scores ranging in ascending order from 1 to 7.

II. Industrial Tests

Variable 16. Teacher's subjective evaluation of hand sewing ability, based on the card entries of the four following sewing variables; accuracy, speed, neatness and handling, variables 17, 18, 19, 20. For the 97 persons of the experiment this correlates with the sum of the gross scores of the four variables, of which it is a subjective summation, to the extent of .662. There was a fair spread of the summation scores.

Variable 21. Teacher's subjective evaluation of power machine operating, based on the card entries of the four following power machine sewing variables; accuracy, speed, neatness, and handling, variables 22, 23, 24, 25. For the 97 this correlates with the sum of the gross scores of the variables, of which it is a subjective summation, to the extent of .881.

Variable 26. Teacher's subjective evaluation of pasting, based on the card entries of the four following pasting variables; accuracy, speed, neatness, handling, variables 27, 28, 29, 30. For the 97 this correlates with the sum of the gross

Digitized by Google

scores of the four variables, of which it is a subjective summation, to the extent of .868.

Variable 31. Grasp of Directions.
A subjective judgment, presumably "ability to imitate" demonstrated performance.

Variable 32. Persistence.
A subjective judgment.

Variable 33. Interest.
A subjective judgment.

Variable 34. Originality.
A subjective judgment, "designing" ability presumably predominating.

Variable 35. Sewing estimate, or "stamp."
The stamps were given numerical scores in ascending order of merit from 0 to 4. The power machine and hand sewing evaluations of variables 16 and 21 are here subjectively combined in the one stamp, making impossible separate evaluation of these two items.

Variable 36. Pasting estimate or "stamp."
The stamps were given numerical scores in ascending order of merit from 0 to 4.

The above descriptions are but meager and inadequate representations of the tests used and the numbers of ratings made. This research is devoted to the evaluation of the qualification card entries, and not to the original ratings from which these are made. As such, the results reported logically fall into two headings:

A. Intercorrelations of various additional tests administered to 97 of one such section of pupils, on whom qualification card records were complete, together with similar intercorrelations of the qualification card data and test scores.

B. Evaluation of the qualification card test scores against average semester school marks in commercial high school subjects of 30 pupils who had passed through this extension school previously, and were found to be attending commercial high school at the time of the investigation.

A. Results of Intercorrelations Based on 97 Pupils

With the aim of investigating the possibilities of standard objective tests in conjunction with the subjective performance tests, additional verbal and non-verbal intelligence tests were given to one of the bi-weekly classes of the extension department. There were 97 girls who had complete test records in all tests.

Digitized by Google

In the results which follow, it is to be remembered that the reliability of the gross scores is somewhat greater than is ordinarily the case with subjective judgments, since each gross score of the qualification card is normally the result of a subjective summation of several personal judgments, made in some cases by two or more teachers independently.

In order to evaluate the qualification card variables against one another, combinations of the scores of the variables were made by adding the gross scores and computing the intercorrelations. These intercorrelations are shown in Table IV.

Of the commercial tests, switchboard, stenography and typewriting (8 + 10 + 11) combined have the highest correlation with the director's estimate (15), $r = .676$. Next in value in predicting the estimate (15) are English and arithmetic combined (5 + 6), $r = .595$; next, general adaptability, personality and appearance combined (12 + 13 + 14), $r = .575$; and finally penmanship and filing (7 + 9), $r = .506$. In view of the magnitude of these correlations, the average of which is .588, it would seem that too much of personal judgment of the director is being injected into the final estimates, over and beyond what the tests themselves show. The weighting of these variables in determining the final "stamp" applied to the card is a subjective matter. The correlations would lead us to believe that much might be gained from using a series of objective weights in determining what is to be the final stamp.

The majority of the industrial tests, as combined, correlate negatively with the commercial estimate. The average of the nine combinations of correlations of industrial tests of the table with the commercial estimate is − .003. The final commercial stamp (15) correlates with sewing estimate to the extent of − .118; and with pasting to the extent of − .008. Taking the tests at their face value, one would conclude that commercial ability and industrial ability in sewing and pasting are correlated slightly negatively. Grasp of directions, persistence, interest and originality in *industrial* work (31 +32 + 33 + 34) combined correlate positively with *commercial* estimate; $r = .306$. This suggests that these subjective tests of character qualities are influenced by the same considerations which make for a correlation of .186 with general adaptability, personality and appearance (12 + 13 + 14) of the commercial tests, and a positive correlation with all other commercial tests. These four subjective judg-

Digitized by Google

ments could well be discarded for a non-verbal intelligence test, for which they are a very inferior substitute.

All industrial tests intercorrelate among themselves positively. The average of nine correlations of the table with the final sewing stamp gives $r = .483$, and with the pasting stamp, $r = .516$.

Combining "accuracy, speed, neatness and handling" by adding the scores, gives lower correlations with the final industrial stamp than the subjective evaluation of these items; in the case of sewing evaluation and sewing stamp, .618 *vs.* .920; sewing estimate and pasting stamp, .543 *vs.* .647; power machine operating and pasting stamp, .718 *vs.* .797. In the other cases the combined score correlates higher with the final stamps; power machine operating and sewing stamp .474 for the estimate *vs.* .569 for the combined score; pasting evaluation with sewing stamp, .305 *vs.* .308, or equal so far as all practical considerations are concerned; and pasting evaluation with pasting stamp, .374 *vs.* .377, again equal for all practical purposes. Inasmuch as the scores in hand sewing have little dispersion, it seems true that subjective bias has entered into these hand sewing scores, and consequently into their final evaluation and is reflected almost perfectly from the evaluation in the sewing stamp applied. Hand sewing evaluation and sewing stamp correlate to the extent of .920. Statistically, the difficulty lies in not making use in the final stamp of the differentiation already secured in the tests. Judgments of "accuracy, speed, neatness, handling" are based on more than one judgment and so have some reliability in themselves. The evaluations can go only by full steps; the summation can add up by partial credits. That is, the evaluation of A's ability must be either 1, 2, or 3; while by summation it may be any score from 4 to 12 inclusive. The correlations afforded by the evaluations of the stamps are such as to prove the feasibility of using the sum of the four.

Alpha (Form 6) and three Thorndike Non-Verbal Intelligence Tests (Forms: I–A, I–L, II–K) were given to the 97 subjects. The intercorrelations are shown in Table V. It is to be noted here, as in the case of all correlations based on these 97 subjects, that while the correlations apparently are low, these correlations are based on a very limited range of talent, the eighth grade of school to which all belonged. Since all are of the same grade, presumably the older pupils are the less intelligent; this is brought out by the four out of five negative intercorrelations. The Toops-Pintner revised direc-

Digitized by Google

TABLE IV

INTERCORRELATIONS OF QUALIFICATION CARD ITEMS. N = 97

(Variable Numbers Correspond to Numbering in the Text)

	Commercial Tests					*Industrial Tests*									
Variable	5 + 6	7 + 9	8 + 10 + 11	12 + 13 + 14	15	16	17 + 18 + 19 + 20	21	22 + 23 + 24 + 25	26	27 + 28 + 29½ + 30	31 + 32 + 33 + 34	35	36	Average *r*, Commercial Tests	Average *r*, Industrial Tests
5 + 6	.....	.508	.435	.270	.595	−.168	−.109	−.010	−.020	.037	.056	.087	−.132	−.044		
7 + 9	.508	.....	.361	.349	.506	−.109	−.135	−.044	−.074	−.038	−.152	.023	−.116	−.117		
8 + 10 + 11	.435	.361	.....	.682	.676	−.142	−.138	.053	−.017	.033	.009	.153	−.132	.000		
12 + 13 + 14	.270	.349	.682	.....	.575	−.178	−.171	−.041	−.207	−.352	−.099	.186	−.196	−.100		
15	.595	.506	.676	.575	.....	−.169	−.053	.010	.041	−.138	.099	.306	−.118	−.008	.588	−.003
16	−.168	−.109	−.142	−.178	−.169	.....	.662	.407	.363	.345	.317	.618	.920	.647		
17 + 18 + 19 + 20	−.109	−.135	−.138	−.171	−.053	.662	.....	.436	.503	.400	.319	.548	.618	.543		
21	−.010	−.044	.053	−.041	.010	.407	.436	.....	.881	.270	.257	.447	.474	.797		
22 + 23 + 24 + 25	−.020	−.074	−.017	−.207	.041	.363	.503	.881	.....	.331	.321	.553	.569	.718		
26	.037	−.038	.033	−.352	−.138	.345	.400	.270	.331	.....	.868	.393	.305	.374		
7 + 28 + 29 + 30	.056	−.152	.009	−.099	−.099	.317	.319	.257	.321	.868	.....	.411	.308	.377		
31 + 32 + 33 + 34	.087	.023	.153	.186	.306	.618	.548	.447	.533	.393	.411	.....	.487	.485		
35	−.132	−.116	−.132	−.196	−.118	.920	.618	.474	.569	.305	.308	.487	.....	.187	−.139	.483
36	−.044	−.117	.000	−.100	−.008	.647	.543	.797	.718	.374	.377	.485	.187	.....	−.054	.516
Distribution Av.	8.35	9.49	14.26	13.78	3.76	1.16	6.64	1.42	7.16	2.37	9.27	9.05	2.27	2.54		
σ	2.425	1.900	3.212	1.725	1.256	.362	.965	.588	1.275	.615	1.542	.963	.682	.704		

Digitized by Google

tions test was also given. This test is a poor measure of whatever abilities are measured by the three non-verbal tests. As a short, easily administered, and perfectly objective, scorable-by-stencil test of intelligence, the test has promise; its correlation with Alpha is higher than the average intercorrelations of the three non-verbal tests. Of the three non-verbal tests, Form I–L, which is most difficult of the three forms, has the highest average intercorrelation. This seems to suggest that more difficult non-verbal tests might prove to be of more value than the present ones.

The correlations of the two intelligence tests and the combined score in the three non-verbal tests, found by adding together the separate gross scores in the three non-verbal tests, are shown in Table VI. The striking contrast between the value of Alpha and the value of non-verbal tests in predicting respectively the commercial test scores and the industrial test scores is shown by the plus sign which indicates which of the two "intelligence" tests is the better predicter of the respective test scores. If we say that the Alpha is

TABLE V

INTERCORRELATION OF MEASURES OF INTELLIGENCE
ALL EIGHTH GRADE PUPILS. $N = 97$

	Intercorrelation With:						
Variable	*Age*	*Directions*	*N. V.–I–A*	*N. V.–I–L*	*N. V.–II–K*	*Alpha*	*Average with other intelligence measures*
Age	....	−.242	−.038	.051	−.111	−.276	−.123
Directions	−.242	....	−.019	.040	.034	.398	.113
N. V.–I–A	−.038	−.019	....	.473	.287	.227	.242
N. V.–I–L	.051	.040	.473	....	.375	.319	.302
N. V.–II–K	−.111	.034	.287	.375	....	.411	.277
Alpha	−.276	.398	.227	.319	.411	....	.339
Variable Average	13.45	15.65	43.02	39.45	45.30	65.80	....
σ	.824	3.395	12.081	9.258	9.519	14.322	....

Digitized by Google

TABLE VI

CORRELATION OF ALPHA AND SUM OF THREE COMBINED SCORES OF NON-VERBAL TESTS WITH VARIOUS TESTS, COMMERCIAL AND INDUSTRIAL. $N = 97$

	Correlation With:													
	English and Arithmetic	*Penmanship and Filing*	*Switchboard, Stenography and Typewriting*	*General Adaptability Personality, Appearance*	*Commercial "Stamp"*	*Hand Sewing. Evaluation of Four Abilities*	*Hand Sewing. Arithmetical Sum of Four Abilities*	*Power Machine Operating. Evaluation of Four Abilities*	*Power Machine Operating. Arithmetical Sum of Four Abilities*	*Pasting. Evaluation of Four Abilities*	*Pasting. Arithmetical Sum of Four Abilities*	*Grasp of Directions. Persistence, Interest and Originality*	*Sewing Stamp*	*Pasting Stamp*
Alpha	.379	.454	.395	.324	.401	.006	−.035	−.113	−.062	.123	.103	.174	−.023	−.044
Three Combined Non-Verbal	.141	.224	.294	.153	.298	.238	.288	.204	.208	.304	.343	.332	.247	.254
Alpha	+	+	+	+	+	.....	.....	.....	.....	.....	.....	.....	.....	.....
Three Combined Non-Verbal	.....	.....	.....	.....	.....	+	+	+	+	+	+	+	+	+

Correlation of Alpha and Three Combined Non-Verbal Tests = .425. In rows 3 and 4, for the respective tests, the plus signs show whether Alpha or the combined non-verbal tests are a better predicter of the given test scores, the plus sign being given to that test which is the better predicter. Without exception, Alpha is a better predicter of all commercial test scores, while the three non-verbal tests combined are in all cases better predicters of the industrial test scores.

Digitized by Google

"typical" of commercial test scores and the non-verbal test of industrial test scores, and if into the subjective evaluation of the four "abilities" in each of the three industrial tests of hand sewing, power machine operating, and pasting there enter largely those personal subjective qualities otherwise designated as "personality," "general adaptability," etc,—then we should expect to find the following:

1. Alpha correlating higher with the subjective evaluation of the four abilities than the objective arithmetical summation of these four abilities. This it does (Table VI) in two cases out of three, the exception being in power machine operating, in which it may be noted, the test product is more easily scored objectively than the hand sewing or pasting products.

2. Non-verbal tests correlating higher with the objective arithmetical summation of the four abilities than with the subjective evaluation. This the combined non-verbal score does in all three cases.

The argument thus seems strongly in favor of scoring the industrial tests solely on the basis of the objective tests, and not allowing for a subjective judgment "over and beyond" what the objective arithmetical summation of the tests alone would give.

B. Follow-up of Students who Entered Commercial High School

From the records of the extension school was found a list of 134 students who had signified their intention of entering the commercial course in Washington Irving High School. From this list only thirty-nine records could be located at the high school (29.1 per cent). This fact is mentioned as showing one of the difficulties encountered in checking up the value of any selection methods. Undoubtedly more than thirty-nine students had entered high school, but had dropped out, since there were a few of the persons who were at the time in their fifth semester at the high school. Our records therefore are lacking in one important element,—knowing the test record and high school record of previous failures. From the thirty-nine records, there were available only the thirty records sufficiently complete in all respects to be used in partial correlation.

The thirty-nine students had been in school for the following lengths of time: 1 semester, 0; 2 semesters, 2; 3 semesters, 4; 4 semesters, 31; 5 semesters, 2; total, 39.

The average per semester percentage school marks were computed in the following subjects: English; mathematics; drawing; stenography; typing; bookeeping.

Digitized by Google

A few unmentioned electives were disregarded in obtaining the average per semester percentage school mark in all subjects. This average school mark is inaccurate through the rating of different persons by different teachers, through changes in the standard of grading from year to year, through unequal lengths of time in school and consequent different numbers of semester courses, together with all the errors inherent in a percentage system of grading. It is a rough criterion of ability to do commercial high school work, but is the best one may obtain without excessive statistical labor spent on only a few cases. Criteria for evaluating test methods are generally subject to such difficulties.

A check on the value of the extension school tests in predicting the probable commercial high school success of these pupils is the correlation between their extension school tests and commercial high school grades in the same or similar work. Table VII shows the results.

It is seen that the typing and stenography tests, approaches to an objective trade test method, are fair predicters of high school marks in those subjects. Extension department arithmetic tests are fair predicters of bookkeeping marks, but not of the marks in algebra,

TABLE VII

TABLE OF CORRELATIONS OF EXTENSION SCHOOL TESTS WITH SAME OR SIMILAR SCHOOL COURSE AVERAGE PER SEMESTER HIGH SCHOOL MARKS

Correlation of	*Number of Cases*	$r =$
School Mark in Typing with Typewriting Test Score	24	.430
School Mark in English with English Test Score	30	.205
School Mark in Stenography with Stenography Test Score	23	.452
School Mark in Mathematics (Algebra) with Arithmetic Test Score	14	+.135
School Mark in Bookkeeping with Arithmetic Test Score	26	.478

Digitized by Google

a more abstract mathematical subject. The English test is a poor predicter of high school marks in English. Probably all these correlations, seemingly not very high, would be higher if the high school group were not of so restricted arrange. Furthermore, the group studied consisted of only the "survivors" and not the "entrants." Again, the extension school estimate must be one which will predict the average success in a subject for as long period of time as five semesters in high school. The results reported above in the case of the English test, a composition test, which in common with most composition tests probably has a high P. E. of the individual score, suggests the need for a revision of this test.

The relationships of the high school commercial course marks to the various "stamps" in commercial tests, sewing tests, pasting tests, are shown in Table VIII. Commercial "stamps" correlate positively to the extent of .403 with commercial high school grades. It is this figure which we hope later to raise by the partial correlation weighting of the various commercial tests. High school marks correlate positively with sewing "stamps" to the extent of .242; and with pasting "stamps" to the extent of .063. There is evidently little relationship between ability to get along well in a commercial high school, as measured, and ability in sewing or pasting, as measured by the

TABLE VIII

INTERCORRELATIONS OF "STAMPS" IN COMMERCIAL, SEWING AND PASTING TESTS WITH AVERAGE PER SEMESTER HIGH SCHOOL MARKS. N=30

Variable	*Correlation with Variable:*			
	1	2	3	4
1	1.000	.403	.242	.063
2	.403	1.000	.346	.336
3	.242	.346	1.000	.597
4	.063	.336	.597	1.000
σ	6.320	1.585	.829	.539
Average	67.20	3.47	2.67	3.10

Variable 1—Average per semester school mark in commercial high school courses.
Variable 2—Estimate "stamp" of commercial ability.
Variable 3—Estimate "stamp" of sewing ability.
Variable 4—Estimate "stamp" of pasting ability.

Digitized by Google

extension school "stamps." The fact that, where we have objective evidence in regard to the pupils' commercial abilities, the correlations of commercial success with the industrial stamps are positive in every case, would lead us to doubt whether the low negative correlations of Table IV really indicate a fundamental negative relationship between "commercial ability" and "industrial ability," or are only the result of the teachers' possible adherence to a "type" philosophy of tests,—the ever popular belief that a person "preeminently" fitted to be a commercial worker will not be a good industrial worker and *vice versa*. Such a chance for bias, not possible with objective trade tests, may impose a very real obstacle to the success of the *subjective* method of judging the results of *objective* tests.

A previous tabulation of the commercial "stamps" and subjective composite of the sewing and pasting "stamps," called "industrial test rating" in the case of 346 successively accumulating qualification cards, yielded a positive correlation between commercial "stamps" and industrial "stamps" of .071. Commercial ability, as measured by these subjective stamp evaluations of commercial tests, is related but slightly to industrial ability similarly determined.

The intercorrelations of the various commercial tests, the correlations with average per semester school marks, the average scores and standard deviations are shown in Table IX.

These tests, excluding the estimate, variable 15, were evaluated by partial correlation yielding the formulae:

In terms of deviations:

$$\frac{x_A}{\sigma_A} = -.015\frac{x_5}{\sigma_5} \text{ (English)} + .173\frac{x_6}{\sigma_6} \text{ (Arithmetic)} + .184\frac{x_7}{\sigma_7} \text{ (Penmanship)} + .036\frac{x_8}{\sigma_8} \text{ (Switchboard)} \tag{8}$$

$$+ .110\frac{x_9}{\sigma_9} \text{ (Filing)} + .120\frac{x_{10}}{\sigma_{10}} \text{ (Stenography)} + .280\frac{x_{11}}{\sigma_{11}} \text{ (Typewriting)} + .137\frac{x_{12}}{\sigma_{12}} \text{ (General Adaptability)}$$

$$- .008\frac{x_{13}}{\sigma_{13}} \text{ (Personality)} - .047\frac{x_{14}}{\sigma_{14}} \text{ (Appearance)}$$

The total correlation of the composite and criterion is, $r_{ic} = .712$.

Digitized by Google

TABLE IX

INTERCORRELATIONS, CORRELATIONS WITH AVERAGE PER SEMESTER SCHOOL GRADES, AVERAGES AND STANDARD DEVIATIONS OF COMMERCIAL TESTS. N=30

		Correlation with Variables:										
	A	5	6	7	8	9	10	11	12	13	14	15
	Average Per Semester School Mark	*English*	*Arithmetic*	*Penmanship*	*Switchboard*	*Filing*	*Stenography*	*Typewriting*	*General Adaptability*	*Personality*	*Appearance*	*Estimate or "Stamp"*
A Average Per Semester Mark	.....	−.015	.425	.455	.117	.140	.436	.509	.569	−.008	.123	.403
5 English	−.015	.....	.105	−.133	.253	.038	.092	.164	.087	−.154	.574	.377
6 Arithmetic	.425	.105	.....	.179	.126	.028	.320	.231	.528	.025	.131	.200
7 Penmanship	.455	−.133	.179	.....	−.046	.013	.436	.187	.481	.091	−.216	−.010
8 Switchboard	.117	.253	.126	−.046	.....	.040	−.173	.203	.120	.360	.292	.209
9 Filing	.140	.038	.028	.013	.040	.....	.045	−.261	.261	.182	.065	.082
10 Stenography	.436	.092	.320	.436	−.173	.045	.....	.336	.427	.298	.001	.673
11 Typewriting	.509	.164	.231	.187	.203	−.261	.336	.....	.440	.434	.267	.521
12 General Adaptability	.569	.087	.528	.481	.120	.261	.427	.440	.....	.112	.148	.430
13 Personality	−.008	−.154	.025	.091	.360	.182	.298	.434	.112	.....	.397	.639
14 Appearance	.123	.574	.131	−.216	.292	.065	.001	.267	.148	.397	.....	.305
15 Estimate or "Stamp"	.403	.377	.200	−.010	.209	.082	.673	.521	.430	.639	.305	.....
Distribution Average	67.20	5.60	.460	5.70	4.77	6.20	4.77	4.30	4.80	6.70	6.30	3.47
σ	6.319	2.375	2.154	2.282	2.232	1.990	2.474	1.847	2.040	1.487	1.847	1.586

Digitized by Google

Or, in terms of gross scores:

$$(9) \quad \bar{x}_A = -.040\,X_5 + .508\,X_6 + .509\,X_7 + .102\,X_8 + .350\,X_9 + .307\,X_{10} + .943\,X_{11} + .424\,X_{12} - .034\,X_{13} - .161\,X_{14} + 53.238.$$

We are now in a position to state the value of the several tests. We shall do this by variable numbers.

Variable 5. English Test. As shown by the very low negative weighting, the test is, for these follow-up cases, worse than useless as now scored. As previously shown, the English test is of almost no value in predicting high school marks in English. A composition test has too high a probable error of the individual score to be of much value as an English test. A Trabue Completion-Test Language Scale, requiring not more than ten minutes, would probably give much better measure of school English marks, and besides is a good intelligence test.

Variable 6. Arithmetic Test. A highly valuable test, exceeded in value only by the penmanship and typewriting tests.

Variable 7. Penmanship Test. Exceeded in value only by the typewriting test.

Variable 8. Switchboard Test. Of little value in predicting high school marks in a commercial course. The test may prove to be of value as a vocational test for switchboard operators. For predicting high school commercial course success, the time spent on this test could well be devoted to other tests of promise, a good intelligence test, for instance.

Variable 9. Filing Test. A desirable test, exceeded in value only by five others.

Variable 10. Stenography Test. A desirable test, about on a par with the filing test.

Variable 11. Typewriting Test. The most valuable test of the group. The spread of scores (standard deviation) is less than in any of the other objective tests: if a greater spread of scores could be obtained, the test might prove to be an even better test than at present.

Variable 12. General Adaptability. This is a subjective judgment test and is about on a par with the filing and stenography tests. The trait probably corresponds most nearly to general intelligence of any of the personal traits judged. In spite of its being a subjective test, it receives a fair positive weighting, sufficient to suggest greater possibilities from an objective intelligence test. The administrative demands upon such an intelligence test are: (1) Short administration time. (2) Short scoring time. (3) High reliability. Some form of omnibus intelligence test will meet these requirements. If the present English test should be abandoned in favor of a Trabue Language Scale, there would be less need of such an intelligence test. Again, the time now spent on the switchboard test would give more than ample time for such an intelligence test, if the switchboard test were abandoned.

Variable 13. Personality. This is another personal judgment test. It receives a negative weighting, which might suggest that "personality" is "bad personality" that remains after a pupil has been judged for intelligence. The time taken for this useless judgment might well be spent in scoring a good intelligence test. The spread of scores is least of all the objective and subjective tests.

Digitized by Google

Variable 14. Appearance. Although the spread of scores is higher than the previous test, it would seem that "personal appearance" is of no concern in commercial high school success! One might, of course, argue that the trait determines to some extent one's success after leaving school and entering business. This seems doubtful, if one may judge from popular opinion, for the homely girl is often preferred in many clerical positions as it is commonly believed that "her chances of being married are less than those of a pretty girl." At any rate, it seems doubtful whether a public school should ever bar anyone from taking a commercial course on such grounds.

The results of this investigation are convincing to the effect that objective tests of the type used can be made a very worthwhile part of the educational machinery involved in preparing children for life's struggles. The results are promising enough to justify the army trade test procedure being adopted in the case of one or two hundred high school freshman students *whose abilities will become known.* It would be but a simple matter to have the same tests, as are now administered to the eighth grade girls, administered to a class of high school students who would attend the extension school for two weeks for the purpose. During this time, an additional number of intelligence tests, reading tests and other tests of promise should be tried. By standardization of the tests on large numbers of pupils who will be certain to attend the high school, we will include the "to-be-failures" in our group as well as the "survivors." This procedure conforms to the procedure of trade tests, adapted to vocational guidance. It should be evident that testing students who have already had extended high school training in stenography, typewriting, and the like, may not be expected *a priori* to yield the desired norms; what is needed is to test a high school freshman class upon entering the high school or soon thereafter. One should also not fail to make the distinction that not all persons in grade school who made a 50 percentile score in the distribution of extension department test scores) will attain that degree of success in high school; rather, if standardized in the manner suggested, "of those people interested enough to enter high school, the tests so standardized will predict such-and-such success, with a given correlation X." Thus, interest, logically kindled through the talks given the extension students by the teachers, may lead the child to make the decision to go to high school, after which the tests will reveal with a rather high degree of accuracy his probable success in the high school course. A correlation of .71 between tests and high school success is undoubtedly

Digitized by Google

TABLE X

ORIGINAL QUALIFICATION CARD AND FOLLOW-UP DATA OF HIGH SCHOOL COMMERCIAL STUDENTS

Person No.	Average Per Semester School Mark	Score in Tests: English	Arithmetic	Penmanship	Switchboard	Filing	Stenography	Typewriting	General Adaptability	Personality	Appearance	Estimate or "Stamp"
1	72	8	8	8	8	8	8	5	8	8	8	5
2	83	8	8	8	5	5	9	8	8	8	5	5
3	68	5	2	8	2	8	9	2	5	5	5	5
4	58	8	2	2	8	5	2	2	2	8	8	2
5	79	8	8	8	5	8	8	5	8	8	8	6
6	73	8	5	5	8	8	2	5	5	8	8	5
7	65	8	5	5	2	2	5	8	5	8	8	5
8	66	5	2	5	5	8	2	2	5	5	8	2
9	67	2	5	8	8	5	5	5	5	8	2	5
10	72	2	5	8	8	5	5	5	5	5	5	2
11	74	2	5	8	2	8	5	5	8	5	5	1
12	66	8	5	8	5	5	2	2	5	5	5	2
13	70	2	5	8	5	8	5	5	5	8	8	5
14	62	8	5	2	5	5	5	5	5	8	8	5
15	64	5	2	8	1	8	5	2	2	8	5	2
16	62	5	8	5	5	8	5	2	2	5	5	2
17	58	5	2	5	2	8	2	2	2	5	2	2
18	64	8	2	2	5	8	2	5	5	8	8	3
19	70	5	5	5	2	8	8	5	8	5	5	5
20	65	5	5	5	5	5	8	5	5	8	8	5
21	53	5	5	5	5	5	2	5	5	8	5	2
22	66	2	2	8	2	2	2	5	5	5	5	2
23	76	5	8	5	5	8	2	5	5	5	8	2
24	73	8	2	5	8	5	5	8	2	8	8	5
25	67	2	8	2	2	2	5	2	5	5	5	2
26	58	8	5	2	5	8	2	2	2	5	8	1
27	65	8	5	8	8	5	8	5	5	8	8	5
28	65	5	2	5	5	5	8	2	2	8	5	4
29	68	2	2	2	5	8	2	5	2	5	5	2
30	67	8	5	8	2	5	5	5	8	8	8	5
Average	67.20	5.60	4.60	5.70	4.77	6.20	4.77	4.30	4.80	6.70	6.30	3.47
σ	6.319	2.375	2.154	2.282	2.232	1.990	2.474	1.847	2.040	1.487	1.847	1.586

Digitized by Google

much better than a consensus judgment of many interviewers without tests would be. Proper additional selection of promising objective tests may better this figure considerably.

One should consider the possibility of rating pupils by their success in industry. This is impossible at present. The most that can be expected of such a vocational guidance department is to make a good prediction of probable high school success.

Table X gives, for reference, the original data used in calculating the correlations of Table IX involving the thirty commercial high school students.

Digitized by Google

CHAPTER IV

GENERAL CONSIDERATIONS IN TESTING AND TRAINING FOR PROFICIENCY AND PROMISE IN THE TRADES

Some of the observations made on the trades, on tradesmen and on trade practices, during the assembly and standardization of ninety army oral trade tests may be of value to anyone who would compare army methods of testing trade proficiency with school methods. The army tests were standardized on men in industry so that the following remarks apply to men in industry and not to soldiers.

Vocational schools may not have been "successful" in the past owing to the fact that they were not, in spite of their claims, aiming to train workers of the kind which they claimed to be training. The vocational schools have claimed to be training tradesmen; they really were attempting to train foremen. Industry has been looking to the trade schools for tradesmen, and necessarily has been disappointed in some cases.

A tradesman's training in industry is intellectually narrow in comparison with the training which the average vocational school would give him. We have already mentioned the specialization which is entering practically all industries and all trades. Trades are being split up into jobs, and jobs into operations, in order that inexperienced and unintelligent persons can quickly become competent and productive, and in order to obtain the profit accruing to the increased production made practicable by a person's being allowed to become highly skilled in one simple operation. No longer does the typical tradesman have to do much figuring for himself. He is "sent out on a job," the standard time for doing which has often been figured out in advance. On the job, a foreman directs his work at every step. If his tools get dull, he hands them in many cases to a toolgrinder; if his belts break, he calls the millwright whose sole business it is to keep the machinery running; if his work is not turning out properly, he may call the toolsetter; and finally, an inspector tests his products.

In consequence of this tendency to specialization, in but few of

Digitized by Google

the trades are there many pertinent items about which trade test questions may be formed. Even in the highly skilled mechanical field of the toolmaker, piece-work wages are being given to some workers.

Again, we have mentioned the use of charts, tables, and mechanical devices used in industry to eliminate arithmetical calculations on the part of the worker. The typical workman is disdainful of studying books. Along with this goes a lack of realization of just how dependent he really is upon the directions of others and upon such book-made aids to trade proficiency. As a result of this attitude, perhaps, he reads trade magazines but occasionally. These remarks do not apply to the foreman, who is not the typical, or journeyman, tradesman. Having but little responsibility for the job, once his toolchest is locked up, the average journeyman's interest in trade matters largely ceases at five P. M. on week days and at twelve o'clock on Saturdays.

But few tradesmen study books on their trades after hours. Wherever, in army standardization of oral trade tests, a boy was found who was pursuing a correspondence course in his trade, he generally made a higher trade test score than other boys of the same length of experience but without the correspondence school course. The general consensus of opinion of foremen in such cases was to the effect that such boys were of more value on the job than boys who did not have the "ambition" to take such courses. Boys who attend night school seem generally superior to boys who do not, but less markedly so than the boys who take correspondence school courses. Possibly one may correctly look upon correspondence school advertisements as an excellent sieve for picking out these boys who have "enough ambition to save up X dollars for self-improvement," whereas the night school appeals to a lesser ability, "enough ambition to go to a free school for self-improvement."

The average journeyman, and the expert to a lesser degree, is limited in the difficulty of trade questions which he is able to answer. It has been found, for instance, that "technical" questions, of the type which one might propound to a superintendent of a machine shop or to a mechanical engineer, were seldom retained for the final examination set in the army oral questions; they were usually discarded because there were so few passes on a question as to make it almost worthless in being able to differentiate between different classes of trade ability. As a striking example, we might cite the case of some

Digitized by Google

forty or fifty millwrights, journeymen and experts, to whom three such "difficult" questions were administered. A millwright is in charge of setting up machinery in a factory and keeping the belts, shafts, etc., in constant repair. One of the easiest of the three questions was: "If a drive pulley, 24 inches in diameter, is running at 200 revolutions per minute, how many revolutions per minute will an 8-inch pulley revolve, which is belted to it?" Only one journeyman and five experts were able to answer the question correctly. Many of them would not even attempt it. The question, it must be observed, is an oral question and some intelligence is needed to keep all the factors in mind at one time; it is not certain that the same results would be secured from using a blue-print of the same situation and asking the question with reference to the blue-print. The question was repeated in its entirety, as often as required by the tradesmen, and yet many of them seemed unable to comprehend the situation clearly enough even to attempt to answer it.

Oral trade test questions must, for tradesmen in the trade, be limited to "practical shop kinks and everyday knowledge." In few trades are there to be found as many as seventy-five questions, of graded difficulty as is necessary, which will fulfill all the requirements of standardization questions; and of these, many will be discarded in the standardization process if rating tests are being made.

The majority of questions must not be technical questions, and on the other hand must be such questions as will sample the daily habits of the workman. For "difficult" questions one must search for the type of question ordinarily regarded as "trade secrets." These trade secrets, when analyzed, prove in most cases to be merely principles of physics, chemistry, general science, or measurement, such as any high school senior should know. Such, for example, is the "rule" that "to get the circumference of an iron ring you multiply the diameter by 3 1/7." It seems at least plausible that the high repute for "skill and accuracy" generally conceded the general machinist, or "mechanic," is due to a misjudgment of the public as to the difficulties involved in measuring to the one-thousandth part of an inch. Using the mechanical measuring equipment which is available in any machine shop, any boy of the intelligence level of a high school graduate should be able to learn to measure to that degree of accuracy in about fifteen minutes. Removing one thousandth of an inch of stock from a piece of iron being turned in a lathe is an even

Digitized by Google

simpler process, with fewer chances for error. "Expert" workmanship, such as is required on but few of the production jobs of industry, means ability to measure to the ten-thousandth part of an inch. For a nominal sum one may now buy sets of "size" blocks, turned out in quantitites, which are "guaranteed not to vary over one hundred-thousandth part of an inch from the stated dimension." Confronted by such statements in the papers, the general public is inclined to overestimate the amount of skill, knowledge, and "trade secrets" possessed by the typical journeyman.

Consistent with the overestimation of trade skill possessed by the typical journeyman, the general public has likewise been inclined to overestimate or misplace the emphasis which should be placed upon intelligence in the acquisition of trade proficiency. A chart, entitled "Occupational Intelligence Standards" published by the Surgeon General's Office, in a pamphlet, "Army Mental Tests," clearly illustrates this fact. The chart shows the range in intelligence covered by the middle 50 per cent of each of the trades indicated. Only a very few workmen, such as laborers, have an average intelligence very decidedly below the average intelligence of the whole population; in perhaps three fourths of the trades of modern industry, "average" intelligence is quite capable of doing work in the manner in which it is now done; in only a very few occupations, such as the clerical vocations and the professions, is an unusually high average amount of intelligence necessary for competency in the trade. If one talks to a foreman who is a leader in his trade, he is rather sure to state the fact that an unusually high degree of intelligence is demanded in order to be a satisfactory worker in *his* trade. The man prides himself on the fact that his trade is a very difficult one to learn, requiring years of patience and effort in order to become a very skilled worker. As a matter of fact, no such condition prevails in at least three fourths of the occupations available to boys entering industry. Such men are prone to overrate the ability required in the present day to learn the trade which they may have learned themselves only by dint of many "hard knocks." Much of the trade information which the tradesman of thirty years ago learned only through experience can now be acquired by a few minutes search in standard reference books on the trade.

Informal tests of tradesmen's ability to recall the questions asked, performed immediately after an examination was finished, have dis-

Digitized by Google

closed the fact that few but the very best and most intelligent tradesmen can remember even half a dozen questions an hour later. The examination should proceed with dispatch from one question to another, thus giving the subject but little time to consider each question and thus fix it in his mind for later recall.

Questions involving complicated mental arithmetic are generally beyond the ability of even the better tradesmen. Questions involving the use of fractions are generally useless in oral trade tests. Mental computation, more elaborate than simple addition of easy numbers is generally too difficult for a tradesman of lower standing than a foreman. Questions involving many limiting conditions, or clauses, are unsatisfactory, owing to the inability of the tradesman mentally to grasp so complicated a situation. By conditions, we mean questions similar to the following: "(1) If you were filing brass, (2) on a speed lathe, (3) using a leather belt, (4) the spindle running at a speed of 1500 revolutions per minute, (5) what.....?" As a practical resulting principle for trade test formulation, a question should not include many conditions if it is to be satisfactory. Essential elements of knowledge may generally be found in such questions, and the question may then be split up into a number of simpler questions. For the same reason, those questions which contain but few words are the surer of being correctly comprehended by the workman. If possible, oral questions should be limited to less than twenty-five words in length, and the shorter the better. Short questions have the additional value of being quickly administered. If one were to use twenty questions of twenty-five words each, or forty questions of ten words each in approximately the same examination time the latter procedure is much more likely to give a selection of tradesmen in the correct order of their abilities. Language ability seems more necessary for a foreman than for an average journeyman; consequently, the longer, more involved questions will be a better measure of foremanship than journeymanship ability. The general education possessed by tradesmen is such as to make one doubtful as to whether success is due to such education, or to intelligence and other qualities of character.

THE EDUCATION OF TRADESMEN

The belief has been fairly common that the tradesman is an "educated" man, and the laborer an "uneducated" man. A tabula-

Digitized by Google

tion, by trades, of the number of school grades completed by 924 tradesmen distributed among thirty trades in the city of Pittsburgh shows that:

1. The median grade at which the various groups of tradesmen left school does not vary greatly, the range being from 7.0 in the case of horseshoers to 10.0 (second year High School) in the case of carburetor man.

2. In the case of six trades, samplings of tradesmen from Pittsburgh and Newark (N. J.) the median grades of the men in the same trades varied in the two cities less than an average of .3 grade.

3. The trades were given a trade rank order depending upon the median grade of education of their tradesmen. The correlation, $\rho = .79$, between the grades and the average trade rank per grade, indicates that, in general, poorly educated men tend to gravitate into trades of which poor education is the more typical, and *vice versa*.

4. Of all the tradesmen, only 26.4 per cent had persisted through a part of high school or more; and 2.8 per cent had completed a part of a college course or more.

5. Experts are generally the best educated of the three classes of tradesmen, apprentices next best, and journeymen least well educated. This seems to indicate that the journeyman stage is a sifting-out stage, possibly a stage of adaptation of mentality and mechanical aptitude or interest to the job, during which time the better educated (highly intelligent) men become experts, foremen, owners, and the like, leaving behind a group of less capable men, "doomed to be journeymen."

6. White laborers, from the Camp Dix personnel office files, were found to be less well educated than colored laborers, while both were very inferior in education to tradesmen.

7. General clerks in New York City were found to have a higher median education than any of the trades.

The educational chart on page 104 (Fig. 29) shows comparable results for the above trades and occupations in regard to education. The results here presented were compiled as the result of interviews with tradesmen in the shop. The length of the bars of this chart indicates the range of the middle 80 per cent in regard to education. The cross-hatched area of the bar indicates the range of the middle 50 per cent. These data are based on a rather small number of cases, the numbers upon which each distribution is based being

Digitized by Google

indicated on the left of the chart. It will be noticed that such trades of this list as are included in the army intelligence chart parallel remarkably the intelligence of those groups. A correlation between the median ranking in education and intelligence of fourteen trades, for which data are available gives $\rho = .78 \pm .10$.

TABLE XI

PERCENTAGE DISTRIBUTION OF MEN, ACCORDING TO NUMBER OF GRADES OF SCHOOL COMPLETED. GENERAL SUMMARY OF 174,857 MEN, VARIOUSLY CLASSIFIED

Grade	*Per Cent Leaving School*					*Per Cent Retained in School to Enter Succeeding Grade:*				
	A	*B*	*C*	*D*	*E*[5]	*A*	*B*	*C*	*D*	*E*
None	2.2		1.0	4.7	0.3	97.8	...	99.0	95.3	99.7
1	0.7		0.3	0.5	[6]	97.1	...	98.7	94.8	99.7
2	1.5	4.8	0.8	0.9	[6]	95.6	...	97.9	93.9	99.7
3	2.8		1.2	1.8	0.2	92.8	...	96.7	92.1	99.5
4	5.0		2.8	3.6	0.5	87.8	95.2	93.9	88.5	99.0
5	8.0	4.0	4.7	3.4	1.9	79.8	91.2	89.2	85.1	97.1
6	11.6	12.7	11.8	7.6	2.9	68.2	78.5	77.4	77.5	94.2
7	10.6	22.3	9.8	10.3	10.0	57.6	56.2	67.6	67.2	84.2
8	31.5	35.0	45.1	34.6	48.3	26.1	21.2	22.5	32.6	35.9
I H. S.	9.7	11.3	6.1	6.0	6.9	16.4	9.9	16.4	26.6	29.0
II H. S.	6.2	6.6	6.4	6.1	8.9	10.2	3.3	10.0	20.5	20.1
III H. S.	2.9	2.2	2.7	4.2	5.1	7.3	1.1	7.3	16.3	15.0
IV H. S.	4.7	1.1	5.0	7.1	10.3	2.6	[2]	2.3	9.2	4.7
A College	0.9	[2]	0.5		1.2	1.7	[2]	1.8	...	3.5
B College	0.6	[2]	0.3		1.1	1.1	[2]	1.5	...	2.4
C College	0.3	[2]	0.4	9.2	0.6	0.8	[2]	1.1	...	1.8
D College	0.7	[2]	1.1		1.7	0.1	[2]	0.0	...	0.1
Post-Grad.	0.1	[2]	0.0		0.1	0.0	0.0	0.0	0.0	0.0
Total	100.0	100.0	100.0	100.0	100.0	*Per Cent with Part H. S. or Better:*				
						26.1	21.2	22.5	32.6	35.9
No. of Men	18,266	120,289[3]	924	30,583[4]	4,795	*Per Cent with Part College or Better:*				
Av. Grade	7.43[1]	7.53	7.83	8.09	8.86	2.6	[2]	2.3	9.2	4.7

[1] Post-graduate college counted as seventeenth grade.

[2] College not tabulated. Hence all percentages slightly too high.

[3] Only 43,358 cards selected at random from the 120,289 were tabulated.

[4] Excludes "aliens, both friendly and enemy."

[5] Samplings only of three groups taken: (*a*) Non-referred men, (*b*) referred and placed, (*c*) referred and not placed. Weighted for column E according to frequency of occurrence: $E = (a \times 17) + (b \times 7) + (c \times 5)$.

[6] Application card did not provide for entry of second and third grades.

Digitized by Google

TABLE XI. KEY TO COLUMNS

A. Unemployed men at Columbus and Dayton, Ohio, Free Employment Offices 1916–1917 (Toops, H. A. and Pintner, R., "Mentality in Its Relation to Elimination from School," *School and Society*, Vol. 7, No. 173 and No. 174, 1918, pp. 507–510 and 534–539).

B. All Working Boys 16, 17 and 18 years of Age subject to Military Duty, residing in villages or towns of 5000 population or over, 1919. Survey of the Military Training Commission, State of New York. Data kindly furnished by Mr. H. G. Burdge, Director of Vocational Training Bureau.

C. Tradesmen, all degrees of skill distributed among 31 trades, Pittsburgh, Pa. Examined in War Department standardization of Army Oral Trade Tests. (Toops, H. A. and Pintner, R., "Educational Differences among Tradesmen," *Journal of Applied Psychology*, Vol. 3, No. 2, 1919, pp. 33–49.)

D. Minneapolis Draftees. "Aliens excluded." Malone, T. J., "When Boys Leave School," *American Review of Reviews*, Vol. 60, No. 6, Dec. 1919, pp. 627–630.)

E. Returned Soldiers, applying for work at the Reemployment Bureau of New York City, 1919. (Toops, H. A., Unsigned news note in the *World*, New York, Sept. 21, 1919.) A somewhat selected group, containing many clerical workers.

Table XI is a distribution table, compiled from all data available to the author, of the grades at which various groups of working men left school. It shows that tradesmen are not, in general, either a superior or an inferior group of people so far as education is concerned. This table serves to emphasize the fact that the typical tradesman is not a high school man. Many of our vocational schools attempt work of high school grade. Where such is done, it obviously will fail to affect the lives of the three fourths of working people who have left school before formal vocational instruction is begun.

We are interested also in the relationship of the amount of education possessed, as measured by the grade at leaving school, to the degree of proficiency in the trade.

THE RELATIONSHIP OF GENERAL EDUCATION TO PROFICIENCY IN A TRADE

Many articles have been written, and much discussion has been devoted to the problem, "Does education pay?" Groups of educators, intent on proving the worth of some specialty in education, or of education in general, have used the argument that such education pays because it may be statistically demonstrated that people with such education receive more wages, rise to higher positions of influence, and the like, than their unlearned brothers. That these argu-

Digitized by Google

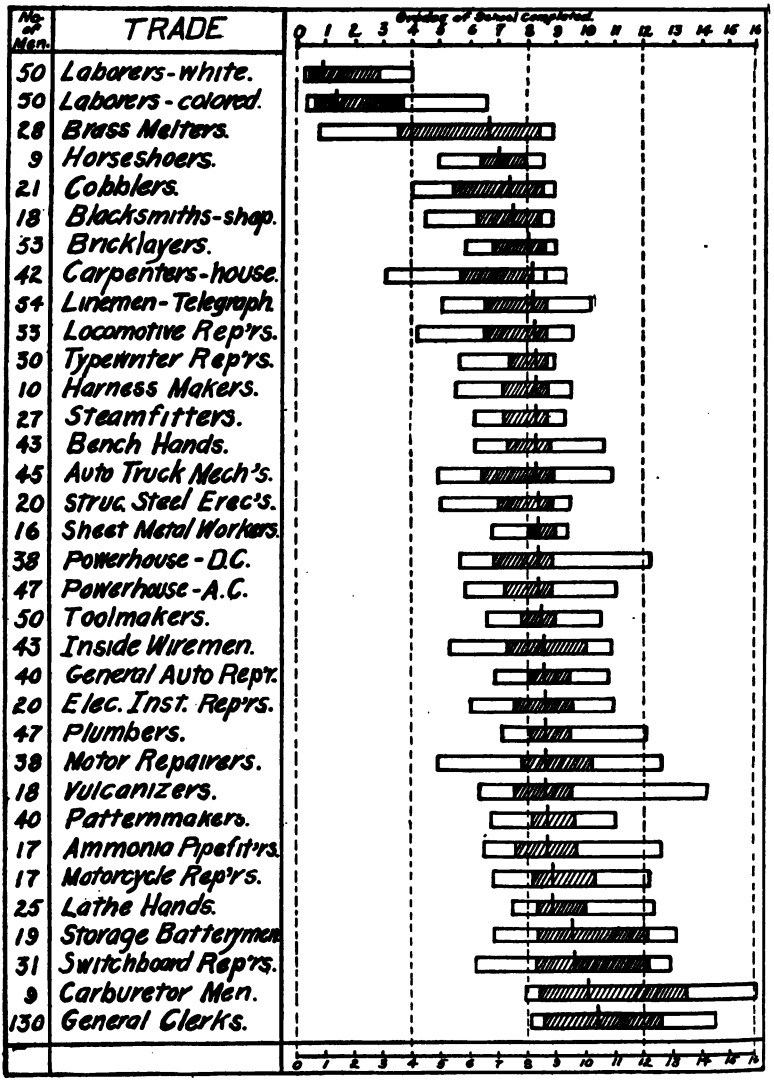


FIG. 29. DISTRIBUTIONS OF GRADE AT LEAVING SCHOOL OF OCCUPATIONAL GROUPS.

The shaded area of each bar shows the range of the middle 50 per cent; the whole bar shows the range of the middle 80 per cent. The vertical line indicates the median of each trade.

Digitized by Google

ments may be fallacious is evident if one but considers the possibility that the group in question may have been a selected group who would have succeeded equally well in industry without the education.

The author was granted permission to evaluate the qualification cards of one hundred and thirty soldiers, just returned from overseas, and applying for jobs as general clerks. Using as a criterion of occupational worth whether or not a man was hired when referred to a prospective employer for a job, and evaluating against this, by partial correlation method, the talents for the job, available from the qualification card, we obtain the following equation:

$$(10)\quad \bar{x}_1 = -100\left(\frac{.496}{.576}\right)x_2 - 20\left(\frac{.496}{2.523}\right)x_3 + 25\left(\frac{.496}{.336}\right)x_4 + 1\left(\frac{.496}{.906}\right)x_5.$$

in which, x_1 is the predicted job value (criterion)
(not hired = score of 1
hired = score of 2)

x_2 is the age in years.
x_3 is the grade at leaving school.
x_4 is the marital condition, (single = score of 1; married = score of 2)
x_5 is the number of dependents.

The composite score, compiled by use of the equation, yields a correlation of .405 with the criterion. The fact that, *other things being equal* in this composite of talents, the more education a prospective employee has the less desirable was he considered by the prospective employer, may be merely a reflection of the economic situation at the time (spring of 1918). Again, one may say that the criterion is too rough for accurate results. The criterion is admittedly rough and defective. The important consideration is that grade at leaving school receives a negative weighting in the composite of talents for the job.

The author was also privileged to secure intensive data on a group of nineteen eyelet machine tenders in a large brass factory. This job is semi-automatic in that the tender merely has to keep a number of machines fed with strip metal, and to watch for defects in the prod-

Digitized by Google

uct, and detect imperfect working of the machine. The following measures of job worth were obtained:

Variable 1. (Criterion). Average daily piece-rate wage over a period of two weeks, reduced to a comparable 10-hour day basis, exclusive of overtime and contingent wage.

Variable 2. Average of three rankings in trade ability of the workers made by the foreman, one week apart, using the slip arrangement order of merit method.

Variable 3. Length of experience on the job to the nearest .01 part of a year.

Variable 4. Oral trade test score. One-word-answer form of question, administered orally and individually by the author.

Variable 5. Grade completed in school.

Using wages as our criterion, it is possible to determine the money value contributed to daily wage by one question answered on a trade test, by one grade in school completed, etc., *all other talents for the job of the composite being constant.*

The formulae resulting are:

In terms of deviations,

$$(11)\quad \frac{\bar{x}_1}{\sigma_1} = .3305\,\frac{x_2}{\sigma_2} + .0890\,\frac{x_3}{\sigma_3} + .5204\,\frac{x_4}{\sigma_4} - .1449\,\frac{x_5}{\sigma_5}$$

Or, in terms of gross measures,

$$(12)\quad \bar{X}_1 = .0508X_2 + .1727X_3 + .0456X_4 - .1408X_5 + K.$$

The correlation of the composite with wages is .687. Again we notice that, although it is an advantage in wages received for a workman to be rated highly by the foreman, to have long experience at the trade, and to receive a high trade test score, grade at leaving school is a disadvantage, *when other talents are equal*, to the extent of fourteen and a fraction cents per day per grade of school attained, in this composite; that is, grade at school receives a negative weighting in spite of its positive correlation with the criterion. When our educators are willing to evaluate their products by the same methods which they would have industry use in evaluating theirs, both will profit by the renewed interest in education. One should not generalize from two isolated instances such as these. The roughness and defectiveness of any criterion by means of which to evaluate the worth of any educational program, any selective method, makes this form of analysis difficult. The method of analysis suggests problems rather than solves them.

Digitized by Google

It is interesting to note, in the first of the above equations, that "trade test ability" is the most important contributor to wages, foreman's judgments of his men next best, while mere experience in years is least important of all four, as judged by the magnitude of the partial correlations.

School surveys, designed to test the needs for vocational schools, have always been made with the point of view that such education would pay. Hence such surveys have literally been searches for facts to "prove the point" that vocational education would pay. This has resulted oftentimes in misplaced emphasis, to the detriment of vocational education. The problem should be "what *kind* of vocational education will pay?" rather than a search for facts to justify vocational education in general.

In such surveys, job analyses are commonly made of a number of trades. The personal requirements, personnel specifications, of workers in the trades (for vocational education has commonly meant trade education) are commonly set forth. These have often been set down, possibly on the basis of a few interviews with foremen, or worse still, on the basis of the opinion of some one or two authorities in vocational education as to what *ought to be* the personal requirements of the worker.

As contrasted with such points of view, we have that developed in the trade-test work of the army, in which the requirements of the worker are any requirements above the abilities of the x-percentile man, his standing in the trade being derived independently of the talents, or tests, for the job,—preferably being some comparative measure of production.

When requirements for the job are viewed in this light, and subjected to statistical treatment, it will probably appear that many of the traits now considered important will prove to have low partial correlations with demonstrated ability on the job.

It is instructive to think of the total sum of any industrial population's talents, *at any one given time*, as being a definitely limited quantity of talent; thus 1,000 men together will have had 8,000 years of school attendance, and so on. When education is considered in this light, almost all large groups of workmen will be found to have a very large percentage of "successful" men who fall far short of the educational standards set up by ordinary job analyses. Table XI gives a summary of the distribution of grade at leaving school of

Digitized by Google

174,857 workmen, as tabulated from the men's own statements recorded on questionnaires. A casual inspection of this table will prove that about half of our working population, with less than eighth grade education, find jobs somewhere. When our vocational educators find vocational education for this half of our population, then such education will more nearly meet the needs of those not "predestined" to success even without the education.

As a partial solution of the problem of vocational placement, we need to recognize traits as being *of variable worth, between definite limits*. The present tendency is to set up merely a minimum limit (usually an "ideal" limit) which, because of a definite limitation of the total sum of human talents, it is ordinarily impossible to maintain. We should be willing to admit that possibly a sixth grade education is a minimum limit for general clerks, that possibly not more than one year of college education is desirable, and that within the aforementioned minimum and maximum limits, a seventh grade education fits better for the job than a sixth grade, an eighth than a seventh, and so on.

This view is at total variance with the economic philosophy which would attempt to educate only the few in order that the world might be well supplied with "workers." Instead, it means conservation of our human resources by reserving to the unintelligent those jobs, any harder than which they are incapable of performing with acceptable speed and accuracy, in short, the highest type of vocational guidance philosophy.

THE USE OF TRADE TESTS AS INCENTIVES TO LEARNING

It is even more difficult for the student of trade subjects to know when he has done his work well than for the academic or scientific student. The academic student, with whom learning is largely content learning, may usually easily classify categorically the amount and quality of his lesson learned as "the amount learned to the point of recall." His test of knowledge and skill gained is the simple test of recall. The scientific student may easily apply the same test to his content matter; the student working on experimental work usually has the self-applicable test of whether his figures agree with the expected figures of the textbook, whether his chemical analysis reveals the same chemical elements which the chemistry instructor put into the test solution, and so on. Trade instruction usually con-

Digitized by Google

tains a minimum amount of content instruction, and much manual manipulation. It is difficult to measure the product of manual manipulation subjectively.

The use of self-administrative performance trade tests for measuring mechanical products may be made of great use as incentives to better trade performance in the case of students sufficiently advanced to apply the tests. Substantially this plan is already in operation in two engineering colleges known to the writer.

In the first of these schools, a simple time standard of performance quality being judged "passable" by the student foreman, has been set up by the best average records of former classes. Those students who beat this time standard are allowed extra credit in grades for the superior performance. There is great rivalry among the boys to turn out product as rapidly as possible, and a keen group coöperative spirit of determination to beat the best record of the previous classes. The method is lacking in the important feature of failing to sufficiently stress "good" work as well as quantity of work, now stressed.

In the second school, percentile norms of performance have been set up. This has obvious advantages over a categorical "standard time," as the student can see for himself not only that he is better or poorer than the standard, but just how much better or poorer. For school purposes the merit of a timed production of work obviously depends upon a measure of dispersion of the distribution of times of former students as well as upon the absolute values of the respective times taken. In industry, the boy who turns out a product in half the time of another is worth twice as much wages as the latter. In vocational school work, the primary emphasis should be upon the acquirement of a certain amount of trade skill, the time required being a matter of secondary importance.

If percentile norms (or index numbers derived from an assumption of a normal curve of distribution of trade ability measured in σ's) were set up for time of performance of all vocational school projects, and if a similar, or preferably an absolute standard of quality of performance were likewise set up for all vocational school projects, then every task performed by the vocational student would be a cumulative test of the degree of his skill at the trade. These tests should be self-administrative. In this simple manner the student may be continually faced by the two generalized measures or objectives of trade work, quality and quantity of product.

Digitized by Google

One of the big advantages of such tests of daily production is the psychological advantage of interest in the work which is created by receiving the rating immediately. A keen interest is usually kindled by any formal examination. If the ratings are not received for several days or weeks, during the time required by the teacher for scoring the papers, much of the original interest in the ratings has been lost by the time the ratings are received. The value of a series of daily grades, determined by the student himself, in keeping him interested in his work can be illustrated by a few percentile ratings which presumably might be received: 75, 83, 70, 86, 83, 20. Were the student required to plot his grades as a daily, or job requirement, the graph afforded by the above percentile ratings would be sufficient incentive to most boys to redouble their efforts on the seventh day. And, inasmuch as the percentiles are based on the work of past classes, the poor pupil has, in effect, an absolute standard to beat and not the "unbeatable" record of relative standing in his class. "Working for grades" under such conditions of objective measurement, free from the instructor's bias, is a highly laudable performance.

For rating quality of product, a series of "limit" or "snap" gages can be easily provided for many products. Where judgments of quality are needed, uniform blanks of analyzed "traits" similar to the army performance test scoring points might be provided and the student be required as part of his task to rate his own product, being graded on his ability to rate his product as well as on the excellence of the product itself.

Formal tests of performance on analyzed important aspects of trade proficiency may also be made self-administrative and self-scorable. The performance measurement test of Chapter I, designed to test the student's ability to use measuring instruments, is scorable by the use of a stencil in but a minute of the student's time. It would seem possible to devise such diagnostic tests of individual weaknesses and strengths in a large number of the operations and processes found in the machine or woodworking shop.

The self-scorable feature of the one-word-answer form of test has been found to have a big appeal to the interest of students in classes in content subjects, where the method has been given a trial. Examinations given weekly or oftener, taking only a half hour or so for the complete examination including the scoring, may be found

Digitized by Google

to be of much more value, both as incentive and as rating method, than much longer examinations given at longer intervals and scorable only by the teacher at a great expense of time and effort. Using such tests, it is not even necessary that the questions be mimeographed. The teacher, standing before the class, merely reads the questions and waits a short time for the students to record their answers before proceeding to the next question. All students may then exchange papers, the answers being scored either right or wrong by the pupils as the teacher slowly reads the correct answer list. The results thus far show that this method may be applied successfully with students as low in ability as the third semester of high school. The method will be found applicable, in some subjects, to the upper grades of the elementary school as well.

Digitized by Google

APPENDIX

STATISTICAL METHODS

The Pearson correlations herein reported, of all variables of n classes with n classes insofar as it was practicable to plot them, were computed by means of stencil and tables of deviation squares by frequencies, using the following specific formula:

$$(13) \quad r = \frac{\frac{N}{2}\left[(\Sigma X^2 + \Sigma Y^2) - \Sigma(X - Y)^2\right] - (\Sigma X)\cdot(\Sigma Y)}{\sqrt{N(\Sigma X^2) - (\Sigma X)^2} \qquad \sqrt{N(\Sigma Y^2) - (\Sigma Y)^2}}.$$

in which X and Y are stencil steps.

For the fourfold Pearson correlations entering into partial correlation regression computations, such cases as the correlation between single—married and not-hired—hired, the formula used is,

$$(14) \quad r_{IV} = \frac{ad - bc}{\sqrt{(a + c)\ (b + d)\ (a + b)\ (c + d)}}$$

in which ad represents the product of the (single—not-hired) by the (married—hired) frequencies.

For the Pearson correlations between a variable of n classes and a categorical variable of two classes, the following formula was used:

$$(15) \quad r = \frac{N\Sigma(b \cdot Y) - B\ (\Sigma Y)}{\sqrt{A \cdot B}\ \sqrt{N(\Sigma Y^2) - (\Sigma Y)^2}}, \text{ in which,}$$

A is the total frequency of the lesser degree of the categories of the X-variable,

B is the total frequency of the greater degree of the categories of the X-variable,

$\Sigma b \cdot Y$ is the sum of the b (greater degree categorical frequencies) by their respective Y-gross scores.

Digitized by Google

For intercorrelations of n variables, using a calculating machine, it has been found very advantageous:

(1) To arrange the crude scores in parallel columns, $X_1, X_2, \ldots Xn$,
(2) subtract from each gross score the lowest gross score of the column yielding the remainders, $X_1 - c_1, X_2 - c_2, \ldots . X_n - c_n$, and
(3) then compute the correlation in parallel columns headed,

$(X_1 - c_1)^2$, $(X_2 - c_2)^2 \ldots . (X_n - c_n)^2$, $(X_1 - c_1) \cdot (X_2 - c_2)$, $(X_1 - c_1) \cdot (X_3 - c_3), \ldots . (X_{n-1} - c_{n-1}) \cdot (X_n - c_n)$.

Most of these multiplications may be done mentally (the squares of all numbers from 1 to 40 may be readily memorized) or by means of a simple multiplication table. The method will be found to be particularly efficient when N is not over 50 persons. The formula used is,

$$r = \frac{N\Sigma XY - \Sigma X \cdot \Sigma Y}{\sqrt{N\Sigma X^2 - (\Sigma X)^2}\sqrt{N\Sigma Y^2 - (\Sigma Y)^2}} \tag{16}$$

The general arbitrary scoring formula for tests given by the multiple choice method, and applicable either to work-limit or time-limit methods of scoring, where n is the number of choices, R the number of attempted questions answered correctly, and W is the number of attempted questions answered wrongly (or omitted), and A is the number of questions attempted, is,

$$S = R - \frac{1}{n - 1} \cdot W. \tag{17}$$

The probable error of an individual score is given by the formula,

$$\text{P. E.}_s = .6745 \frac{(n)}{(n - 1)} \sqrt{\frac{R \cdot W}{A}}. \tag{18}$$

If a criterion is available, so as to use the partial regression equation of

(19) $S = R + C \cdot W$, wherein C has the value determined by Thurstone,[1]

[1] Thurstone, L. L., "A Scoring Method for Mental Tests." *Psych. Bul.*, Vol. 16, No. 7, 1919, pp. 235–240.

Digitized by Google

$$C = \frac{\sigma_R \ (r_{IR} \cdot r_{RW} - r_{IW})}{\sigma_W \ (r_{IW} \cdot r_{RW} - r_{IR})} \tag{20}$$

Then the generalized formula for the probable error of an individual score is,

$$\text{P. E.}_s = .6745 \left| (C - 1) \right| \sqrt{\frac{R \cdot W}{A}} \tag{21}$$

The formula for determining the reliability of n forms of a test with n other forms of a test the reliability of one form with a second being r_{11}, is,

$$r_{nn} = \frac{n \cdot r_{11}}{1 + (n - 1) r_{11}}. \tag{22}$$

This equation may be solved for n, yielding,

$$n = \frac{r_{nn} \ (1 - r_{11})}{r_{11} \ (1 - r_{nn})} \tag{23}$$

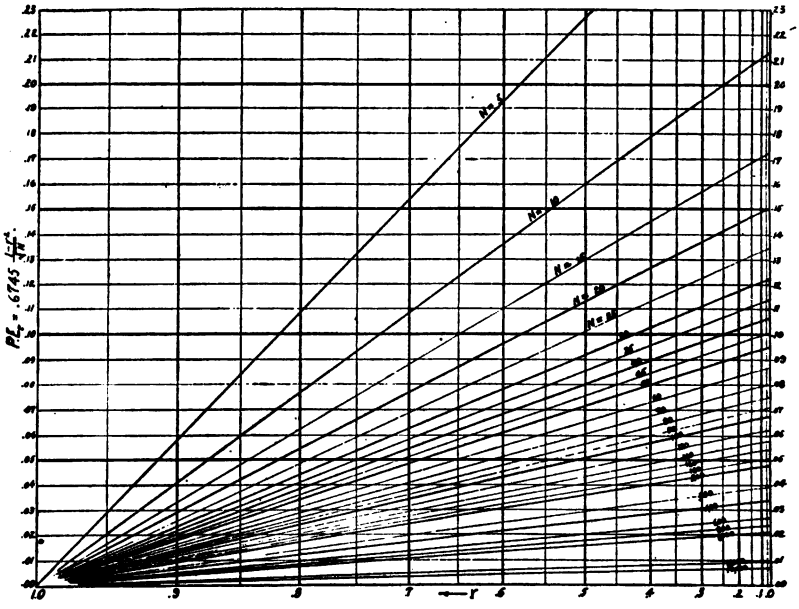


FIG. 30. CHART FOR FINDING PROBABLE ERRORS OF PEARSON *r*'s.

Digitized by Google

Probable Errors of Correlation Coefficients

For the convenience of those who may be interested in the P. E.'s of the correlation coefficients herein given, the chart of Fig. 30 is published. Such straight line charts may be adapted to solving many equations commonly used in routine test work.

To use the chart: Locate the r, of which the P. E. is desired, along the lower base line. Project this vertically by an imaginary line until intersecting the diagonal line which represents the number of cases on which the correlation coefficent is based. Then, project this intersection either to the right or left hand margin of the chart where the P. E.$_r$ may be read directly. Interpolation must be done for values of N not on the chart, and for values of r not on the .05-point division lines. Even with the interpolation, the accuracy is as great or greater than is commonly required.

Digitized by Google

BIBLIOGRAPHY

A Selected List of Fifty References Arranged Alphabetically

1. Bawden, W. T. *The Army Trade Tests.* Department of the Interior, Bureau of Education, Circular No. 4, Washington, April, 1919. 28 pp.
2. Bingham, W. V. *Measuring a Workman's Skill: The Use of Trade Tests in the Army and Industrial Establishments.* Reprinted from Bulletin No. 30, National Society for Vocational Education Proceedings of the St. Louis Convention, 1919. 12 pp.
3. Bloomfield, M. *Youth, School and Vocation.* Boston, Houghton Mifflin Co., 1915.
4. Brewer, J. M. *The Vocational Guidance Movement.* New York, Macmillan Co., 1918.
5. Bridges, J. W. The Relation of Intelligence to Occupation as Indicated by Army Data. (An abstract of a paper presented before the American Psychological Association, Baltimore, Christmas, 1918.) *Psychological Bulletin,* Vol. 16, No. 2, 1919, pp. 42–43.
6. Burtt, H. E. Employment Psychology in the Rubber Industry. *Journal of Applied Psychology,* Vol. 4, No. 1, 1920, pp. 1–17.
7. Chapman, J. C., and Toops, H. A. A Written Trade Test: Multiple Choice Method. *Journal of Applied Psychology,* Vol. 3, No. 4, 1919, pp. 358–365.
8. Chapman, J. C. Mental Tests in Industry. *Personnel* (National Association of Employment Manager's Bulletin), Vol. 1, No. 3, 1919, pp. 1 and 9.
9. Chapman, J. C. The Measurement of Physics Information. *School Review,* Vol. 27, 1919, 748–756.
10. Cody, S. *Commercial Tests and How to Use Them.* New York, World Book Co., 1919, 216 pp.
11. Coss, John J., and Outhwaite, L. *Personnel Management, Topical Outline and Bibliography.* Classification Division, Adjutant General's Office, Washington, January, 1919, pp. 58.
12. Dean, A. D. *The Worker and the State.* New York, The Century Co., 1910, 355 pp.
13. Federal Board for Vocational Education. Employment Management Series, Bulletins 1–9, 1919–1920. (Nine pamphlets.)
14. Fish, E. H. Principles of Employing Labor. *Industrial Management,* Vol. 57, No. 4, 1919, pp. 277–281.
15. Hollingworth, H. L. *Vocational Psychology.* New York, D. Appleton and Co., 1916, 308 pp.
16. Jaques, M. P. Mental Tests for Typists and Stenographers. *Industrial Management,* Vol. 58, No. 2, 1919, pp. 145–147.
17. Johnson, J. F. Testing Applicants for Factory Positions. *Industrial Management,* Vol. 58, No. 2, 1919, pp. 109–110.

Digitized by Google

18. JONES, E. D. *Bibliography of Employment Management.* Federal Board for Vocational Education, Employment Management Series, Bulletin No. 9, Washington, 1920.
19. JONES, E. R. Personnel Activities from the Point of View of the Individual Operative. *Personnel* (official organ of the National Association of Employment Managers), Vol. 1, No. 8, 1919, pp. 10–11.
20. KELLEY, T. L. Principles Underlying the Classification of Men. *Journal of Applied Psychology*, Vol. 3, No. 1, 1919, pp. 50–67.
21. KELLY, R. W. Dangers in Rating Employees. *Industrial Management*, Vol. 58, No. 1, 1919, pp. 35–42.
22. KELLY, R. W. *Hiring the Worker.* New York, The Engineering Magazine Co., 1918, 250 pp.
23. KELLY, R. W. Selecting and Training Interviewers. *Industrial Management*, Vol. 57, No. 4, 1919, pp. 263–270.
24. LAMB, J. P. Intelligence Tests in Industry. *Industrial Management*, Vol. 58, No. 1, 1919, pp. 21–23.
25. LINK, H. C. *Employment Psychology.* New York, Macmillan Co., 1919, 440 pp.
26. MARCUS, L. Vocational Selection for Specialized Tasks. *Journal of Applied Psychology*, Vol. IV, Nos. 2 and 3, 1920, pp. 186–201.
27. MEINE, F. *Job Specifications.* Federal Board for Vocational Education, Employment Management Series No. 3, Bulletin 45, 1919.
28. OTIS, A. S. The Selection of Mill Workers with Mental Tests. *Journal of Applied Psychology*, Vol. IV, No. 4, 1920, pp. 339–341.
29. REILLY, P. J. *The Selection and Placement of Employees.* Federal Board for Vocational Education, Employment Management Series No. 2, Bulletin No. 49, Washington, 1919, 82 pp.
30. ROBINSON, E. S. The Analysis of Trade Ability. *Journal of Applied Psychology*, Vol. 3, No. 4, 1919, pp. 352–357.
31. RUML, B. The Extension of Selective Tests to Industry. *The Annals of the American Academy of Political Science*, Vol. 81, whole No. 170, January, 1919, pp. 38–46.
32. SCOTT, W. D. Changes in Some of Our Conceptions and Practices of Personnel. *Psychological Review*, Vol. 27, No. 2, 1920, pp. 81–94.
33. SLICHTER, S. H. *The Turnover of Factory Labor.* New York, D. Appleton and Co., 1919, 460 pp.
34. SNEDDEN, D. *Vocational Education.* New York, The MacMillan Co., 1920, 587 pp.
35. TEAD, O., and GREGG, B. *Outline of Job Analysis.* Valentine & Gregg, New York, 1918, 19 pp.
36. TEAD, O. *The Labor Audit.* Federal Board for Vocational Education, Employment Management Series No. 8, Bulletin 43, 1920.
37. TOOPS, H. A., and PINTNER, R. Educational Differences Among Tradesmen. *Journal of Applied Psychology*, Vol. 3, 1919, pp. 33–49.
38. TOOPS, H. A., and PINTNER, R. Mental Tests of Unemployed Men. *Journal of Applied Psychology*, Vol. 1, No. 4, pp. 325–341; Vol. 2, No. 1, 1918, pp. 15–25.

Digitized by Google

39. TOOPS, H. A., and PINTNER, R. Variability of the Education of Unemployed Men. *The Journal of Applied Psychology*, Vol. 2, 1918, pp. 207–218.
40. THORNDIKE, E. L. Fundamental Theorems in Judging Men. *Journal of Applied Psychology*, Vol. 2, No. 1, 1918, pp. 67–76.
41. U. S. Employment Service Bulletin. Model Demonstration Office Opened in New York City for Skilled Labor, Vol. 2, No. 2, Feb. 14, 1919, p. 3.
42. U. S. Department of Labor, Bureau of Labor Statistics. Descriptions of Occupations, Washington, 1919 (8 pamphlets).
43. War Department, Adjutant General's Office. The Making of a Trade Test. *Personnel*, Vol. 1, No. 7, 1918, pp. 25–28.
44. War Department, Adjutant General's Office. *Personnel Work in the U. S. Army*. Washington, January, 1919 (a pamphlet).
45. War Department, Adjutant General's Office. *The Right Man in the Right Place in the Army*. Washington, 1919, 68 pp.
46. War Department, Committee on Classification of Personnel. *The Personnel System of the U. S. Army*. Washington, 1919 (in two volumes: Vol 1, History of the Personnel System, Vol. 2, The Personnel Manual).
47. War Department, Committee on Education and Special Training. *Final Report of the National Army Training Detachments*. (S. A. T. C.) War Department, Washington, 1919, 179 pp.
48. War Department, Division of Psychology, Surgeon General's Office. *Army Mental Tests*, Washington, 1918, 23 pp. (a pamphlet).
49. WATSON, M. S. Trade Test Principles in Industry. *Industrial Management*, Vol. 59, No. 3, 1920, pp. 247–250.
50. WARDNER, C. A. Applying the Army Trade Tests to Vocational Schools. *Industrial Arts Magazine*, Vol. 8, 1919, pp. 402–403.

Digitized by Google

VITA

Herbert Anderson Toops was born at Kiousville, Ohio, September 18, 1895. After six years of attendance at country schools he entered Midway High School, Sedalia, Ohio, graduating therefrom in 1912. From 1912–1914 he attended Ohio Wesleyan University; and from 1914 to 1917, Ohio State University in the College of Engineering, and later in the College of Education. He received the degrees of B. A. and B.Sc. in Education in 1916 and M. A. in 1917.

During the war, he entered the Trade Test Division of the Committee on Classification of Personnel in the Army, where he served in the various capacities of standardizer, assembler and statistician on oral trade tests until January, 1919. At that time he entered the newly formed Technical Interview Methods Section of the United States Department of Labor for the purpose of introducing army trade test methods into the U. S. Employment Service. Later, he served as statistical assistant in the statistical evaluation of the National Research Council group intelligence tests. During 1919–1920 he was a research scholar at Teachers College, Columbia University. Since August, 1920, he has been engaged as specialist in the construction of vocational tests for the E. and R. Schools of the Army.

His publications, to date, in full or in joint authorship, are as follows:

1. PINTNER, R., and TOOPS, H. A. A Chart for Rapid Computation of Point Scale Scores. *Journal of Delinquency*, Vol. 2, No. 4, 1917, pp. 209–210 plus chart.
2. PINTNER, R., and TOOPS, H. A. A Mental Survey of the Population of a Workhouse. *Journal of Delinquency*, Vol. 2, No. 5, 1917, pp. 278–287.
3. PINTNER, R., and TOOPS, H. A. Mental Tests of Unemployed Men. *Journal of Applied Psychology*, Part 1, Vol. 1, No. 4, 1917, pp. 325–341; Part 2, Vol. 2, No. 1, 1918, pp. 15–25.
4. TOOPS, H. A., and PINTNER, R. Mentality in its Relation to Elimination from School. *School and Society*, Vol. 7, No. 174, 1918, pp. 507–510; and No. 175, 1918, pp. 534–539.
5. PINTNER, R., and TOOPS, H. A. A Revised Directions Test. *Journal of Educational Psychology*, Vol. 9, No. 3, 1918, pp. 123–142.
6. PINTNER, R., and TOOPS, H. A. A Drawing Completion Test. *Journal of Applied Psychology*, Vol. 2, No. 2, 1918, pp. 164–173.

Digitized by Google

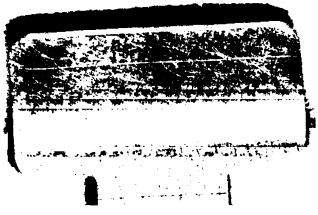
7. TOOPS, H. A., and PINTNER, R. Variability of the Education of Unemployed Men. *Journal of Applied Psychology*, Vol. 2, No. 3, 1918, pp. 207–218.
8. TOOPS, H. A., and PINTNER, R. A Chart for the Determination of I. Q. Values. *Journal of Delinquency*, Vol. 3, No. 6, 1918, p. 272 plus chart.
9. TOOPS, H. A., and PINTNER, R. Educational Differences Among Tradesmen. *Journal of Applied Psychology*, Vol. 3, No. 1, 1919, pp. 33-49.
10. TOOPS, H. A. Plotting Equations of Three Variables in Mental Measurements. *Psychological Review*, Vol. 26, No. 4, 1919, pp. 317–326.
11. TOOPS, H. A., and PINTNER, R. Mentality and School Progress. *Journal of Educational Psychology*, Vol. 10, No. 5–6, 1919, pp. 253–262.
12. CHAPMAN, J. C., and TOOPS, H. A. A Written Trade Test: Multiple Choice Method. *Journal of Applied Psychology*, Vol. 3, No. 4, 1919, pp. 358–365.
13. TOOPS, H. A., and PINTNER, R. Curves of Growth of Intelligence. *Journal of Experimental Psychology*, Vol. 3, No. 3, 1920, pp. 231–242.

Digitized by Google

Digitized by Google

Digitized by Google

YC 04604



Digitized by Google

Digitized by Google